DOCUMENT RESUME

ED 465 287 FL 027 338

AUTHOR Thompson, Lynn E.; Kenyon, Dorry M.; Rhodes, Nancy C.
TITLE A Validation Study of the Student Oral Proficiency Assessment (SOPA).
INSTITUTION Center for Applied Linguistics, Washington, DC.; Iowa State Univ. of Science and Technology, Ames. National K-12 Foreign Language Resource Center.
SPONS AGENCY Center for International Education (ED), Washington, DC.
PUB DATE 2002-05-00
NOTE 71p.
CONTRACT P229A3005
PUB TYPE Reports - Research (143) -- Tests/Questionnaires (160)
EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Elementary Education; Elementary School Students; FLES; Immersion Programs; *Language Proficiency; *Language Tests; *Oral Language; Second Language Instruction; *Test Validity; Testing
IDENTIFIERS Oral Proficiency Testing

ABSTRACT

This study validated the Student Oral Proficiency Assessment (SOPA), an oral proficiency instrument designed for students in elementary foreign language programs. Elementary students who were tested with the SOPA were also administered other instruments designed to measure proficiency. These instruments included the Stanford Foreign Language Oral Skills Evaluation Matrix (FLOSEM) and the CAL Student Self-Assessment (SSA). Testing sites involved students in either a FLES program, a content-enriched FLES program, a foreign language experience, a two-way partial immersion program, or a total immersion program. Students and teachers at each site completed background questionnaires to gather program information and ethnographic data. Results indicated that the SOPA measured proficiency as intended. The correlation between SOPA levels and FLOSEM ratings were strong, while correlations between SOPA levels and the SSA were relatively low (possibly due to differences in mode of assessment and rating procedures). Overall, the content-enriched FLES program and the partial immersion programs, which had the most variation among students, provided the strongest empirical correlations between the SOPA and the other instruments. (Contains 39 tables and 20 references.) (SM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 465 287

A Validation Study of the Student Oral Proficiency Assessment (SOPA)

Lynn E. Thompson
Dorry M. Kenyon
Nancy C. Rhodes

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Donna Christian
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1

May, 2002
Final Version

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Center for Applied Linguistics, Washington, DC
Iowa State University National K-12 Foreign Language Resource Center
Ames, IA

This study was conducted with funding from the U.S. Department of Education, Office of Postsecondary Education, Center for International Education, under grant no. P229A3005 to Iowa State University

FL 027 338

BEST COPY AVAILABLE

A Validation Study of the Student Oral Proficiency Assessment (SOPA)
Table of Contents

I. Introduction 1
Background 1
Validation of SOPA 3

II. History and Background of SOPA 5
SOPA Interview Description 5
SOPA Rating 6

III. Overview of Study Design 7

IV. Methodology – Study 1: FLES SOPA Validation
Subjects 8
Less Intensive Foreign Language Programs (Sites A-E) 8
French
Site A 8
Site B 9
Site C 9
Spanish
Site D 9
Chinese
Site E 9
Instrumentation 11
SOPA 11
Student Self-Assessment (SSA) 11
FLOSEM 12
Background Questionnaire 12
Procedures 12
Data Analysis 13
Feedback to the Sites 13
FLES Validation Study 13

V. Results – Study 1: FLES SOPA Validation 15
Subset A: Chinese 15
SOPA Ratings and Student Self-Assessment Scores 15
SOPA Levels and Teachers' FLOSEM Ratings 16
Subset B: Spanish 17
SOPA Ratings and Student Self-Assessment Scores 17
SOPA Levels and Teachers' FLOSEM Ratings 18
Subset C: French 20
SOPA Ratings and Student Self-Assessment Scores 20
SOPA Levels and Teachers' FLOSEM Ratings 20
French Teacher A 21
French Teacher B 21

French Teacher C 22
French Teacher D 23
Summary of Results 24
SOPA Levels and Student Self-Assessment Scores 24
SOPA Levels and FLOSEM Ratings 24
Conclusion 25

VI. Methodology – Study 2: Immersion SOPA Validation 26
Subjects 26
Total Immersion – Site F 26
Two-Way Partial Immersion – Site G 26
Instrumentation 28
Procedures 28
Data Analysis 28
Feedback to the Sites 28
Immersion Validation Study 28
Background Variables Considered 30

VII. Results – Study 2: Immersion SOPA Validation 31
Subset F: Total Immersion
SOPA Ratings and Student Self-Assessment Scores 31
SOPA Levels and Teachers' FLOSEM Ratings 32
Subset G: Two-Way Partial Immersion
SOPA Ratings and Student Self-Assessment Scores 35
SOPA Levels and Teachers' FLOSEM Ratings 37
Summary of Results
SOPA Levels and Student Self-Assessment Scores 40
SOPA Levels and Teachers' FLOSEM Ratings 40

VIII. Methodology – Study 3: Inter-rater Reliability 41

IX. Results of Data Analysis for Inter-rater Reliability Study 42
Results of Correlations Across Common Examinees 42
Results of Correlations Between Pairs of Raters 42
Summary of Results 43
Conclusions and Recommendations 44

X. Summary, Conclusions, and Future Directions
Summary 45
Conclusions 46
Future Directions 46

References 48

Charts

Chart A: Language Programs Participating in SOPA Validation Study: FLES PROGRAMS 10
Chart B: Language Programs Participating in SOPA Validation Study: IMMERSION PROGRAMS 27

Tables

Table 1: SOPA Listening Comprehension Levels and Student Self-Assessment Mean Totals for Chinese 15
Table 2: SOPA Fluency Levels and Student Self-Assessment Mean Totals for Chinese 15
Table 3: SOPA Listening Comprehension Levels and FLOSEM Comprehension Mean Ratings for Chinese 16
Table 4: SOPA Fluency Levels and FLOSEM Fluency, Vocabulary, Pronunciation, Grammar and Fluency Total Means for Chinese 17
Table 5: SOPA Listening Comprehension Levels and Student Self-Assessment Mean Totals for Spanish 18
Table 6: SOPA Fluency Levels and Student Self-Assessment Mean Totals for Spanish 18
Table 7: SOPA Listening Comprehension Levels and FLOSEM Comprehension Mean Ratings for Spanish 19
Table 8: SOPA Fluency Levels and FLOSEM Fluency, Vocabulary, Pronunciation, Grammar, and Fluency Total Means for Spanish 19
Table 9: SOPA Listening Comprehension Levels and Student Self-Assessment Mean Totals for French 20
Table 10: SOPA Fluency Levels and Student Self-Assessment Mean Totals for French 20
Table 11: SOPA Comprehension Levels and FLOSEM Comprehension Means (French Teacher A) 21
Table 12: SOPA Fluency Levels and FLOSEM Fluency, Vocabulary, Pronunciation, Grammar and Fluency Total Means (French Teacher A) 21
Table 13: SOPA Comprehension Levels and FLOSEM Comprehension Means (French Teacher B) 22
Table 14: SOPA Fluency Levels and FLOSEM Fluency, Vocabulary, Pronunciation, Grammar, and Fluency Total Means for French (French Teacher B) 22
Table 15: SOPA Comprehension Levels and FLOSEM Comprehension Means (French Teacher C) 22
Table 16: SOPA Fluency Levels and FLOSEM Fluency, Vocabulary, Pronunciation, Grammar and Fluency Total Means (French Teacher C) 23
Table 17: SOPA Comprehension Levels and FLOSEM Comprehension Means (French Teacher D) 23
Table 18: SOPA Fluency Levels and FLOSEM Fluency, Vocabulary, Pronunciation, and Grammar Means (French Teacher D) 24
Table 19: Comparison SOPA Listening Comprehension Levels and Student Self-Assessment Totals for Total Immersion 31
Table 20: Comparison of SOPA Fluency Levels and Student Self-

Assessment Totals for Total Immersion 31
Table 21: Comparison of SOPA Grammar Levels and Student Self-Assessment Totals for Total Immersion 32
Table 22: Comparison SOPA Vocabulary Levels and Student Self-Assessment Totals for Total Immersion 32
Table 23: Comparison of SOPA Listening Comprehension Levels and FLOSEM Listening Comprehension Ratings for Total Immersion 33
Table 24: Comparison of SOPA Fluency Levels and FLOSEM Fluency Ratings for Total Immersion 33
Table 25: Comparison of SOPA Fluency Levels and FLOSEM Pronunciation Ratings for Total Immersion 33
Table 26: Comparison of SOPA Grammar Levels and FLOSEM Grammar Ratings for Total Immersion 34
Table 27: Comparison of SOPA Vocabulary Levels and FLOSEM Vocabulary Ratings for Total Immersion 34
Table 28: Correlations Between SOPA Levels and FLOSEM Ratings for Total Immersion 35
Table 29: Comparison of SOPA Listening Comprehension Levels and Student Self-Assessment Totals for Two-Way Immersion 35
Table 30: Comparison of SOPA Fluency Levels and Student Self-Assessment Totals for Two-Way Immersion 36
Table 31: Comparison of SOPA Grammar Levels and Student Self-Assessment Totals for Two-Way Immersion 36
Table 32: Comparison of SOPA Vocabulary Levels and Student Self-Assessment Totals for Two-Way Immersion 37
Table 33: Comparison of SOPA Comprehension Levels and FLOSEM Comprehension Ratings for Two-Way Immersion 37
Table 34: Comparison of SOPA Fluency Levels and FLOSEM Fluency Ratings for Two-Way Immersion 38
Table 35: Comparison of SOPA Grammar Levels and FLOSEM Grammar Ratings for Two-Way Immersion 38
Table 36: Comparison of SOPA Vocabulary Levels and FLOSEM Vocabulary Ratings for Two-Way Immersion 38
Table 37: Correlations Between SOPA Levels and FLOSEM Ratings For Two-Way Immersion 39
Table 38: Correlations Across Common Examinees 42
Table 39: Correlations Between Pairs of SOPA Raters 43

Appendix A
Student Oral Proficiency Assessment (SOPA) Background Information Questionnaire
Appendix B
CAL Student Self-Assessment for French (FLES Version)
CAL Student Self-Assessment for Language (Two-Way Immersion Version)

A Validation Study of the Student Oral Proficiency Assessment (SOPA)

Center for Applied Linguistics
Iowa State University National K-12 Foreign Language Resource Center
Lynn E. Thompson, Dorry M. Kenyon, Nancy C. Rhodes

ABSTRACT

This study, conducted by the Center for Applied Linguistics in conjunction with the Iowa State University National K–12 Foreign Language Resource Center, addresses the validation of the Student Oral Proficiency Assessment (SOPA), an oral proficiency instrument designed specifically for children in elementary school foreign language programs. Validity refers to whether the SOPA measures "proficiency" as it is intended. In order to validate the SOPA, other instruments that aim to measure proficiency were administered to students who were also tested with the SOPA. The Stanford Foreign Language Oral Skills Evaluation Matrix (FLOSEM) and the CAL Student Self-Assessment (SSA) were selected as the instruments to be administered to students at seven sites representing the broad spectrum of elementary foreign language programs.

Sites represented the following program models: foreign language experience (30 minutes/week or 30 minutes/week plus 1-hour of music taught in the language); FLES (30 minutes/day, 2 days a week or 30 minutes/day, 3 days a week); content-enriched FLES (50 minutes/day, 5 days a week, with subject matter including physical education, art, and language arts); two-way partial immersion (50% of daily instruction in target language); and total immersion (nearly 100% of instruction in target language). Languages of instruction included Chinese, French, and Spanish. A background questionnaire was used to gather program information and ethnographic data on the student and teacher population at each site to ensure proper interpretation of assessment results.

The results of the validation study (correlations and comparisons of means) found evidence that the SOPA does appear to measure proficiency as intended. The correlation between SOPA levels and FLOSEM ratings were strong (students with higher SOPA levels received higher FLOSEM ratings), while correlations between SOPA levels and the SSA were relatively low (possibly due to differences in mode of assessment and rating procedures). Overall, the two participating programs (content–enriched FLES and partial immersion) that provided the strongest empirical correlations between the SOPA and other instruments were those that had the most variation among students (testing cohorts included students who were in a range of grade levels and/or included native vs. non-native speakers of target language).

Three recommendations are made for future studies. First, more research needs to be conducted on the inter-rater reliability of the instrument, a critical element of any oral language assessment. Second, a follow-up study is recommended that includes sites with a wider range of variation among students. Finally, research recently conducted as part of program evaluations on the use of the SSA and the SOPA should be formally analyzed and published.

A VALIDATION STUDY OF THE STUDENT ORAL PROFICIENCY ASSESSMENT (SOPA)

I. Introduction

With the dramatic increase in the number of elementary schools offering foreign language instruction in the last 2 decades, there has been increased interest in finding better ways to evaluate the language proficiency of young students. As elementary schools increasingly focus on accountability and standards in all subject areas, the more the language profession needs to be able to accurately demonstrate how well students are doing in their foreign language classes. In addition, professional agreement on the importance of early-start, long-sequence language programs (ACTFL, 1998), along with the recommendations of the *Standards for Foreign Language Learning* (1996), reinforce the importance of assessing the effects of early language instruction. According to Donato (1998), "Failure to do so may result in serious damage to the future health and credibility of elementary foreign language programs nationwide and in increased marginalization relative to other subject areas" (p. 170). The major obstacle in this effort of accurately assessing students' language is that there are few, if any, validated language assessment instruments that are geared toward assessing children in the communicative fashion in which they have been taught.

The Iowa State University's National K-12 Foreign Language Resource Center, in conjunction with the Center for Applied Linguistics, addressed this issue by revising and validating one of the few oral proficiency instruments that is designed specifically for children in elementary school language programs: the Student Oral Proficiency Assessment (SOPA). The overall project included revising the SOPA rating scale using the national foreign language standards, performance guidelines for K–12 learners, and immersion benchmarks; developing a less–intensive version of the SOPA (also referred to as the FLES version) as well as revising the immersion version; developing an administrator's manual; and conducting a validity study on both versions of the SOPA. This paper will focus on one aspect of the project—the process of validation of the SOPA.

Background

There is a dearth of research on the use of oral proficiency language tests for children in the United States because only recently have school districts and researchers been investigating the issues for this age group. (Outside of the

United States, however, there is a body of research on early immersion programs in Canada as well as on assessing foreign language ability in primary schools in Australia.) A few of the key references that help put this study in context are described below.

Donato (1998) provides a comprehensive overview of the critical issues in assessing the early language learner. He laments the use of adult-based constructs of proficiency for children and recommends the collection of longitudinal empirical data showing what young learners actually know and are able to do in various instructional models and at particular points during the learning process. He raises questions that the profession needs to address, such as these: What knowledge should ultimately inform test construction and alternative forms of assessment? What communicative abilities and cultural knowledge need to be assessed in the early language learner? What assessment procedures are best suited for young learners? Donato and colleagues G. R. Tucker and J. L. Antonek have identified some of the complexities of assessing linguistic and cultural gains of children. The issues include the identification of foreign language abilities that describe a proficient early language learner and discriminate proficiency in a foreign language across an array of oral and literate skills, subskills, and tasks.

Another key reference on assessing the foreign language ability of young learners is the newly released *ACTFL Performance Guidelines for K-12 Learners* (1998) that expanded upon the *ACTFL Proficiency Guidelines* (1986) by focusing on foreign language use by students in elementary through high school language programs. These guidelines are the performance standards (that define how well students perform) designed to accompany the national content standards (that define what students learn). These much-awaited guidelines are designed to help foreign language educators recognize language performance across levels of proficiency, modes of communication, and criteria for accuracy.

A search for research studies specifically on the validation of oral proficiency instruments for young children results in even fewer citations because of the informal nature of classroom-based assessments that are currently being developed and used by teachers. The few tests that have been validated are standardized exams that were normed on bilingual education students (mostly in the 1970s and 1980s) and thus not necessarily appropriate for the target audience of elementary school foreign language students. One useful source of information on assessment

instruments currently being used in elementary and middle schools is L. Thompson's *Foreign Language Assessment in Grades K-8: An Annotated Bibliography of Assessment Instruments* (1997). The bibliography provides a thorough review of both traditional and alternative foreign language assessment instruments.

The instruments featured are used in a wide variety of program models, ranging from those offering language instruction 75 minutes a day to those offering total immersion instruction almost 100% of the day. Validity and reliability information was requested from each test contributor, and when provided, is included in the "test development and technical information" section of the test description. Many of the tests, as expected, do not have such information because they were developed by teachers for classroom use and were not developed using rigorous test development procedures. The introduction to the volume explains that, "For classroom-based assessment, data on reliability and validity is neither available nor necessary given the orientation and purposes of the assessment. The more weight and importance given to the decisions that are based on assessment instrument results, the more important validity and reliability become" (p. xxii). That said, this is the most comprehensive review of tests available, and it provides a realistic snapshot of the technical rigor of assessment instruments currently in use in schools.

Other current references on early language assessment include those that address specific program evaluations (e.g., Donato, Antonek, & Tucker, 1994; Heining-Boynton, 1990; Lipton, 1996; Tucker, Donato, & Antonek, 1996), descriptions of specific instruments (Educational Testing Service, 1993; Lapkin, Argue, & Foley, 1992), or general guidelines for assessment (Clementi & Sandrock, 1994; Genesee & Upshur, 1996; Rhodes, Rosenbusch, & Thompson, 1996; TESOL, 2001).

Validation of SOPA

The purpose of studying the validity of an instrument is to investigate whether the instrument measures what it is supposed to measure. Technically speaking, validity is "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment" (Messick, 1989, p. 13).

For validation of this assessment, the research question was the following: Does the SOPA appear to measure "proficiency" as it is intended? Information from the SOPA was gathered for descriptive and research purposes. In other contexts, another question would be asked: Is the SOPA valid for the decisions that are being made with the test? In that case, appropriateness of decisions made on student progress, program evaluation, or effectiveness of teaching would be discussed.

How is proficiency defined for the SOPA? The SOPA is designed "to assess students' ability to understand and speak a foreign language" in a global manner. The SOPA seeks to capture what students know and can do in a foreign language, both in the classroom environment and beyond. In order to validate whether the SOPA measures this construct, other instrument(s) that also purport to measure proficiency need to be administered to the students who are tested with the SOPA. For this study, the Stanford Foreign Language Oral Skills Evaluation Matrix (FLOSEM) (Padilla, 1994) and the CAL Student Self-Assessment (SSA) (1996) were used.

II. History and Background of SOPA

The purpose of the Student Oral Proficiency Assessment (SOPA) is to determine students' oral proficiency and listening comprehension in a foreign language. The SOPA was developed in response to requests from school districts for an alternative language assessment instrument for students in the lower elementary grades. The instrument is based on the CAL Oral Proficiency Exam (COPE), an interactive immersion assessment developed for fifth and sixth graders in response to a need for "an oral interview-type test that would elicit normal speech and would yield global ratings of proficiency" (Rhodes & Thompson, 1990). Although it is based on a reliable test that has been validated, the SOPA itself had never been validated and has not been formally packaged for distribution to teachers.

The SOPA was developed in 1991 to evaluate students in the first grade Spanish partial immersion program at Woodland Elementary School in Oak Ridge, Tennessee. Since then, the instrument has been adapted for various grade levels and various programs types and is now being used to assess children in Grades 1–5. Recently, the instrument was adapted for use in non-immersion French, German, Japanese, and Spanish elementary school language programs. The SOPA has been used in a variety of programs around the country, including Alexandria (Virginia) Public Schools Spanish immersion program; Arlington (Virginia) Public Schools two-way Spanish partial immersion programs; Calvert County (Maryland) Public Schools Japanese immersion program; Foreign Language Immersion and Cultural Studies School (Detroit, Michigan); Georgia Elementary School Foreign Language Model programs; and Metropolitan School District of Lawrence Township French and Spanish immersion programs (Indianapolis, Indiana). (See Boyson, Rhodes, & Thompson, 1998; Rhodes, 1988; and Rhodes, 1998 for articles on the SOPA.)

SOPA Interview Description

The SOPA is designed for the language and developmental levels of children in partial and total immersion programs (including two-way immersion) and FLES programs. It consists of four parts that are set in an interview format: (1) *listening comprehension* (students are asked to point to, identify colors of, and respond to commands using various fruit manipulatives); (2) *informal questions and TPR commands* (students are asked to respond to personal questions about their age, family, pets, etc., and respond to commands with appropriate actions); (3)

science and language usage (for immersion students only) (students show knowledge of science concepts by describing a series of four pictures) *OR listening and speaking activities* (students manipulate colorform objects and people in a dollhouse) *(for FLES students only);* and (4) *story telling (for immersion students only)* (students are given a picture book and are asked to tell the story in Spanish by describing what's happening in the pictures) *OR describing a classroom scene (FLES students only)* (students respond to questions and commands about classroom objects and activities); and (5) *supporting an opinion/persuasion* (students support their views on school rules) (*for immersion students only*). Two students at a time are assessed by two examiners in a non-stressful, friendly environment. The interview takes approximately 10-15 minutes to complete. The goal of the assessment is to show what the students *can* do with language, not what they cannot do.

The test aims to get the students to use as much language as possible in a short period so that there will be a large body of data on which to base the ratings. The rating and interviewing tasks are divided between two examiners: one rater and one interviewer. This ensures that the interviewer can focus entirely on guiding the students to their highest possible level of performance in both listening comprehension and oral fluency, and the rater can focus on rating the students objectively and accurately. The SOPA is conducted entirely in the target language. Ideally, the SOPA should not be used as the only assessment of a student's progress in proficiency development, but should be used in conjunction with teacher observations and other evaluations of the student's daily oral and written work.

SOPA Rating

Students' language is rated holistically. Students are evaluated in pairs to facilitate dialogue between each other and between them and the examiners. The SOPA rating scale uses the first six levels of a nine-level scale from the COPE test, which is based on the proficiency guidelines of the American Council on the Teaching of Foreign Languages. SOPA students receive one of six ratings for comprehension and fluency (whereas the COPE ratings for fifth and sixth graders include comprehension, fluency, vocabulary, and grammar). However, in recent years, a revised version of the longer nine-level COPE scale has been used successfully with immersion students on the SOPA, providing more detailed information for the raters.

The six levels of the SOPA rating scale are Junior Novice-Low, Junior Novice-Mid, Junior Novice-High, Junior Intermediate-Low, Junior Intermediate-Mid, and Junior Intermediate-High. The comprehension ratings range from "recognizes a few familiar questions and commands" (Junior Novice Low) to "usually understands speech at normal speed, though some slow-downs are necessary; can request clarification verbally" (Junior Intermediate High). The

fluency ratings range from "conversations are limited to an exchange of memorized sentences or phrases" (Junior Novice Low) to "maintains conversation with remarkable fluency but performance may be uneven; uses language creatively to initiate and sustain talk" (Junior Intermediate High). The longer COPE scale includes three higher levels: Junior Advanced, Junior Advanced-High, and Superior. In addition, this extended scale provides the option of rating student performance in terms of two additional criteria: vocabulary and grammar. When the SOPA is given annually, a student's ratings are expected to increase gradually, revealing his or her progress in the foreign language.

III. Overview of Study Design

The results of this research will be presented as three separate studies because of the multiple versions of the instrument involved. The studies will be presented as follows: (1) the validation study of the FLES version of the SOPA; (2) the validation study of the immersion version of the SOPA; and (3) a small-scale inter-rater reliability study of SOPA raters involved in the first study.

IV. Methodology – Study 1: FLES SOPA Validation

Subjects

The SOPA validation study required administration of the SOPA to students in programs representative of the broad spectrum of foreign language instructional models available in American public elementary schools. Nineteen teachers or foreign language supervisors representing a variety of programs were identified and trained in the use of the SOPA by CAL staff during a week long *Performance Assessment Institute* at the Iowa State University National K-12 Foreign Language Resource Center in June 1997. From this group, sites were chosen representing less intensive instruction (foreign language exploratory [FLEX], FLEX and music, foreign language in the elementary school [FLES] twice–a–week, and FLES three–times–a–week), content-enriched (five–times–a–week) foreign language, and total immersion. An additional two-way partial immersion site, whose staff was proficient in the use of the SOPA, was added as the seventh site so that all major program models were represented. This selection provided five different sites for administration of the elementary foreign language SOPA (see Chart A) and two sites for administration of the immersion SOPA.

All students in the study had been enrolled in their school's language program either since kindergarten/grade 1 or, in the case of one site, since the starting grade for foreign language instruction at their school (3rd grade). There were four students in the latter group that had only been in their school's program for one year. Brief descriptions of each site follow. Sites are grouped according to program type and language and are identified as A, B, C, D, E, F, and G. Note that immersion sites (F and G) are described in study 2.

<u>Less Intensive Foreign Language Programs (Sites A-E)</u>

French

Students were tested in French at three different sites. For purposes of this validation study, their results have been combined except as noted in the data analysis and results sections of this report.

Site A

Thirty fifth graders with "high academic potential" were selected for testing at this site, which offers French instruction for 30 minutes a week, Grades K-6. The program is located in a south–central state and has enjoyed strong support from parents and administration. Students have benefited from having the same teacher throughout the program. Unlike students at the other sites visited, these students had some advance preparation for their SOPA interview in the form of practice and review of vocabulary and functions associated with the tasks in the SOPA.

Site B

This program is located in the same city as Site A. Students in this Grade 3-6 French program have the unique opportunity to receive French instruction 30 minutes a week and music taught in French 1 hour a week with the same teacher. The students were randomly selected from a magnet school population. With the exception of four students, all students have been in the program for 2 years. Many of the students at the school are considered gifted and talented. Twenty fourth graders and four third graders were interviewed with the SOPA. This program receives strong support from the administration and the community.

Site C

This long-established, 30-minute twice-a-week French program is located in a small city in the Southeast. French instruction is available to all students from Grades 1-5. Sixty fifth graders who have studied French since first grade were randomly selected to participate. Students were selected from two different schools (each with a different teacher) in the same city. Students have had a number of different instructors. Both fifth-grade teachers are American but fluent in French.

Spanish

Site D

Thirty third graders and thirty fourth graders from two different schools within the same program were randomly selected to participate. They have received instruction in Spanish for 30 minutes, three times a week with the same teacher since first grade. This site is located in the Southwest. The program was established through a federal Foreign Language Assistance Program grant.

Chinese

Site E

This program offers instruction in Chinese for 50 minutes a day, five times a week in Grades K-5. Students in Grades 2-5 were selected for participation in the SOPA validation study based on their availability at the time of the site visit. They had all been in the program since kindergarten. Students benefit from native-speaking instructors and the presence of native-speaking students in each class. The program is called "content-enriched" because instruction is enriched through art and physical education activities in Chinese as well as language arts. This site is located in the Midwest.

Chart A: Language Programs Participating in SOPA Validation Study
FLES PROGRAMS

Site Location	Program Type	Student Selection	Grade	# of Students per Grade	Native Speakers of Target Language (Students)	Teachers' Language	Special Characteristics
Chinese							
Midwest Site E	Content–enriched FLES 5 x 50 minutes per week K-5	Whole program, based on students available at time	5 4 3 2	6 6 16 12	17% (1) 33% (2) 6% (1) 25% (3)	Native (2)	Students had a different teacher for kindergarten. Teach PE and Art and Language Arts in Chinese.
Spanish							
Southwest Site D	FLES 30 minutes 3 x week (2 schools) 1-6	Full range (random)	4 3	30 30	0%	Non-native (1)	Students have had the same teacher since Grade 1.
French							
Southeast Site C	FLES 30 minutes 2 x week (2 schools, 2 teachers) 1-5	Full range (random)	5 5	30 30	0%	Non-native (2)	Students have had a number of different teachers since kindergarten.
South–Central Site B	FLEX 30 minutes + music in French (1 hour) 4-6	Random, but from higher achievers	4 3	20 4	0%	Non-native (1)	Students have had the same teacher since Grade 3. Teacher is fluent; teaches the music class in French.
South–Central Site A	FLEX 30 minutes a week K-6	Random, but from higher achievers	5	30	0%	Non-native (1)	Students have had the same teacher since kindergarten.

Instrumentation

The validation study plan required the administration of at least one or possibly two other instruments at the same time as the SOPA for comparison purposes. In addition, a background questionnaire was developed (see Appendix A) to provide information useful to the interpretation of assessment results on all instruments. A search of available foreign language assessment instruments revealed a lack of similar oral proficiency assessments that could be used in both less intensive and immersion programs. The Foreign Language Oral Skills Evaluation Matrix (FLOSEM), developed by Dr. Amado Padilla of Stanford University, was selected for use since it was found to be the closest match to the two versions of the SOPA. The FLOSEM is a matrix used by teachers to rate individual students' speaking, listening comprehension, pronunciation, grammar, and vocabulary skills in the foreign language. Like the SOPA, the FLOSEM views student proficiency in terms of what the student can do and provides a descriptive rating matrix that reflects the natural progression of acquisition of foreign language skills.

SOPA

There are two versions of the SOPA—the immersion version and the less intensive version. The SOPA assesses two students at a time in a friendly, informal manner. A pair of students is given a number of tasks ranging from identifying fruits, to answering personal questions, to describing a picture or series of pictures, to retelling a story. At the conclusion of the interview, the students are assigned a rating using a rating matrix adapted from the Proficiency Guidelines of the American Council on the Teaching of Foreign Languages (ACTFL).

Student Self-Assessment (SSA)

Given the growing interest in and appreciation of student self-evaluation, a student self-assessment questionnaire, keyed to the content and levels of difficulty of the SOPA, was developed as a companion assessment (see Appendix B). A preliminary version for less intensive foreign language instruction programs was piloted with 40 students in a twice-a-week elementary French program. In addition to subjecting their responses to item analysis, students' comments on the instrument (i.e., clarity of instructions, items, and rating scale) were gathered and factored into the revision process. In addition, two immersion versions of the Student Self-Assessment, one in the target language alone and one in both English and the target language, were developed with direct input from both partial and total immersion teachers for use in Study 2.

FLOSEM

The FLOSEM was designed for use with less intensive foreign language programs as well as immersion programs, Grades K-12 or higher. The FLOSEM is a 6-level teacher observation matrix for fluency, listening comprehension, grammar, vocabulary, and pronunciation. For each skill, there are six possible levels ranging from the equivalent of a very beginning speaker to a native speaker of the target language. The instrument has been used extensively with students in programs for both commonly and less commonly taught foreign languages. To date, formal validation studies have not been published on the FLOSEM, but an informal study in which students and their instructors rated their proficiency using the FLOSEM showed a very high correlation between the two scores (A.M. Padilla, personal communication, Fall 1996).

Background Questionnaire

A background questionnaire was developed for gathering program information and ethnographic data on the student and teacher population at each site (see Appendix A). The questionnaire was based on a number of previously developed and administered CAL questionnaires used for information gathering in elementary foreign language programs in prior research. This questionnaire was administered to ensure proper classification of the sites and interpretation of assessment results.

Procedures

Interviewers included foreign language educators trained at the Performance Assessment Institute in the use of the SOPA and two foreign language educators at the partial immersion site where prior versions of the SOPA had been used. These educators also aided CAL staff in the collection of background data on the program and facilitated the administration of the Student Self-Assessment and the FLOSEM. CAL staff or consultants who had been involved in the development of the different versions of the SOPA and/or trained in its use served as raters at the sites. There was an average of 42 students per site.

The assessment instruments were administered in the same order at each site. Prior to SOPA administration, students completed the student self-assessment. A background questionnaire was sent to the principal and/or foreign language supervisor at each site, and CAL staff reviewed their responses during the site visit. At almost all sites, SOPA interviews were conducted and rated over a 2-day period. The students' classroom teachers completed the FLOSEM either immediately prior to the SOPA testing if the classroom teacher was serving as the SOPA interviewer, or following the testing if he/she was not serving as interviewer.

Data Analysis

Feedback to the Sites

For each site, student averages for all instruments were calculated, discussed, and interpreted in a written report to the school. For programs where students were drawn from more than one school or came from differing language backgrounds, comparative information was included in the report. Statistically significant differences in student performance were also reported and interpreted. These reports are not included here because they identify each site by name and report results according to site rather than by language or program type.

FLES Validation Study

For purposes of FLES SOPA validation, the data were handled differently. The emphasis in the validation study was not on reporting average student performance on each of the instruments for each site, but on looking at the relationship between student performance on the SOPA, the SSA and the FLOSEM and seeing if this information confirms the validity of the SOPA.

If the SOPA is a valid assessment of language proficiency, then relationships between SOPA outcomes and outcomes on the SSA and FLOSEM are expected. All instruments attempt to capture an assessment of oral language skills. The SSA captures information directly from the language learner. The FLOSEM captures assessment information based on the observations of the student's language teacher. The SOPA captures assessment information in a more formal assessment situation. It seeks to provide a valid "snapshot" of the student's oral language skills, captured in an efficient manner. Clearly, if the SOPA is valid, outcomes on the SOPA should be positively correlated to outcomes on the other, less formal assessments.

For the SSA, correlational studies were conducted to address the following hypotheses:

1. The higher the SSA score, the higher the SOPA comprehension score should be.
2. The higher the SSA score, the higher the SOPA fluency score should be.

To test the first hypothesis, SSA scores were correlated with SOPA comprehension scores. To confirm the second hypothesis, SSA scores were correlated with SOPA fluency scores.

For the FLOSEM, the analysis involves not only fluency and listening comprehension but also pronunciation, grammar, and vocabulary (seen in terms of fluency), components that are not treated separately in the SOPA.
The following hypotheses were examined for the FLOSEM:

1. The higher the FLOSEM comprehension score, the higher the SOPA comprehension score should be.
2. The higher the FLOSEM fluency score, the higher the SOPA fluency score should be.
3. The higher the FLOSEM grammar score, the higher the SOPA fluency score should be.
4. The higher the FLOSEM vocabulary score, the higher the SOPA fluency score should be.
5. The higher the FLOSEM pronunciation score, the higher the SOPA fluency score should be.
6. The higher the FLOSEM fluency total score, the higher the SOPA fluency score should be.

To confirm the first hypothesis, FLOSEM comprehension and SOPA comprehension were correlated. For the second, FLOSEM fluency and SOPA fluency were correlated. For the third, FLOSEM grammar and SOPA fluency scores were correlated. For the fourth, FLOSEM vocabulary and SOPA fluency scores were correlated. For the fifth, FLOSEM pronunciation scores and SOPA fluency scores were correlated. Finally, for the sixth, FLOSEM fluency, pronunciation, grammar, and vocabulary scores were combined to make a "total" fluency score and compared to SOPA fluency.

Unlike the SOPA, which was rated by CAL staff or consultants who had received intensive training together, each classroom teacher at each site rated the FLOSEM. These teachers were trained separately by different CAL consultants. It was anticipated that teacher ratings on the FLOSEM would vary considerably. For the Chinese and Spanish sites, this was not the case. In the case of the French data, which was drawn from three different sites, FLOSEM ratings did vary considerably. Therefore, comparisons were made by teacher rather than by the entire cohort of students for French.

V. Results – Study 1: FLES SOPA Validation

Results of data analysis are presented by language subsets: Chinese, Spanish, and French. Correlations between the SOPA ratings and the Student Self-Assessment total scores, the SOPA ratings and the FLOSEM ratings (comprehension, fluency, vocabulary, pronunciation, and fluency component totals), along with a comparison of mean scores on all instruments, are presented and discussed.

Subset A: Chinese

SOPA Ratings and Student Self-Assessment Scores

For Chinese, SOPA listening comprehension ratings ranged from Jr. Novice Mid to Jr. Intermediate High and SSA totals from 21.40 to 28.50 (see Table 1). SOPA oral fluency ratings ranged from Jr. Novice Low to Jr. Intermediate High. Corresponding SSA scores ranged from 19.00 to 29.50 (see Table 2).

Table 1: SOPA Listening Comprehension Levels and Student Self-Assessment Mean Totals for Chinese

SOPA Listening Comprehension Levels	Number of Students Per Level	Student-Self Assessment	
		Mean	Standard Deviation
Jr. Novice Mid	21	21.40	2.64
Jr. Novice High	6	22.17	4.87
Jr. Intermediate Low	4	22.50	3.42
Jr. Intermediate Mid	3	27.33	2.52
Jr. Intermediate High	4	28.50	2.38

Table 2: SOPA Fluency Levels and Student Self-Assessment Mean Totals for Chinese

SOPA Fluency Levels	Number of Students Per Level	Student-Self Assessment	
		Mean	Standard Deviation
Jr. Novice Low	2	19.00	1.41
Jr. Novice Mid	20	21.40	2.76
Jr. Novice High	8	23.12	4.12
Jr. Intermediate Mid	5	27.40	2.51
Jr. Intermediate High	2	29.50	0.71

The correlation between SOPA comprehension ratings and Student Self-Assessment total scores was .50 (sig., p = .002) for Chinese. The correlation between SOPA oral fluency ratings and SSA total scores was .58 (sig., p = .000).

Both correlations between the SOPA ratings and SSA total scores and comparisons of SOPA ratings and corresponding mean total SSA scores support the hypothesis that the two instruments are assessing the same thing. That is, the higher the SOPA rating (listening comprehension or oral fluency) the higher the total score on the SSA.

SOPA Levels and Teachers' FLOSEM Ratings

Table 3 shows the SOPA listening comprehension levels achieved, the number of students who achieved each level, and their mean FLOSEM comprehension rating. Below each mean rating, in parenthesis, is the standard deviation of that mean. Table 3 illustrates, without exception, the higher the student SOPA listening comprehension level was, the higher the FLOSEM comprehension level. SOPA listening comprehension levels ranged from Jr. Novice Mid to Jr. Intermediate High. Corresponding FLOSEM comprehension mean ratings climbed steadily from 1.10 to 5.40.

Table 3: SOPA Listening Comprehension Levels and FLOSEM Comprehension Mean Ratings for Chinese

SOPA Listening Comprehension Levels	Number of Students	FLOSEM Comprehension
Jr. Novice Mid	20	1.10 (0.26)
Jr. Novice High	6	1.17 (0.26)
Jr. Intermediate Low	4	2.00 (0.71)
Jr. Intermediate Mid	3	3.00 (1.00)
Jr. Intermediate High	5	5.40 (0.89)

Table 4 illustrates that the higher the SOPA oral fluency level was, the higher the FLOSEM fluency rating. SOPA oral fluency levels ranged from Jr. Novice Low to Jr. Intermediate High. FLOSEM fluency ratings ranged from 1.00 to 6.00. Table 4 also provides mean FLOSEM ratings for vocabulary, pronunciation, grammar, and total fluency (a composite score comprised of fluency, vocabulary, pronunciation, and grammar). Comparisons between SOPA oral fluency levels and FLOSEM vocabulary and FLOSEM grammar show the same trend as FLOSEM fluency. As SOPA oral fluency levels increase, so do mean scores for vocabulary and grammar. FLOSEM pronunciation mean ratings showed some slight variation from this trend. On the average, Jr. Novice High level students (n = 8) were rated slightly higher (2.13) than the one Jr. Intermediate Low level student (2.00). In light of the differing numbers of students at each SOPA level and the standard deviation for FLOSEM pronunciation ratings for Jr. Novice High level students (0.89), this difference is not considered important.

Table 4: SOPA Fluency Levels and FLOSEM Fluency, Vocabulary, Pronunciation, Grammar, and Fluency Total Means for Chinese

SOPA Fluency Levels	Number of Students	FLOSEM Fluency	FLOSEM Vocabulary	FLOSEM Pronunciation	FLOSEM Grammar	FLOSEM Fluency Total
Jr. Novice Low	2	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	5.00 (0.00)
Jr. Novice Mid	19	1.18 (0.30)	1.08 (0.25)	1.68 (0.67)	1.08 (0.25)	6.13 (1.54)
Jr. Novice High	8	1.62 (0.74)	1.44 (0.68)	2.13 (0.83)	1.63 (0.74)	8.38 (3.45)
Jr. Intermediate Low	1	2.00 (0.00)	1.50 (0.00)	2.00 (0.00)	2.00 (0.00)	9.00 (0.00)
Jr. Intermediate Mid	6	4.00 (1.41)	3.92 (1.56)	4.20 (1.17)	4.00 (1.41)	20.08 (6.93)
Jr. Intermediate High	2	6.00 (0.00)	6.00 (0.00)	6.00 (0.00)	6.00 (0.00)	30.00 (0.00)

The correlation between SOPA listening comprehension levels and FLOSEM comprehension was .81 (sig., p = .000) for Chinese. The correlation between Chinese SOPA oral fluency levels and FLOSEM fluency ratings was .75 (sig., p = .000). Correlations between Chinese SOPA oral fluency levels and other FLOSEM components (all considered part of fluency) were as follows: vocabulary .79 (sig., p = .000), pronunciation .71 (sig., p = .000), and grammar .81 (sig., p = .000).

Thus, in the Chinese study, high and significant correlations between SOPA levels and FLOSEM ratings and similar trends in SOPA levels and FLOSEM mean ratings were found. This outcome provides strong evidence for the validity of the SOPA as an assessment of oral language skills. The hypothesis was upheld that the higher the SOPA listening comprehension or oral fluency rating, the higher the corresponding (composite or individual) FLOSEM rating.

Subset B: Spanish

SOPA Ratings and Student Self-Assessment Scores

For Spanish, SOPA listening comprehension ratings ranged from Jr. Novice Low to Jr. Novice High. SSA totals ranged from 22.20 to 23.38 (see Table 5). SOPA oral fluency ratings ranged from Jr. Novice Low to Jr. Novice High, and SSA total scores went from 22.04 to 23.91 (see Table 6). It is important to note, however, that only one student was at the Jr. Novice High level, while 34 students were at the Junior Novice Mid level. The standard deviation for SSA scores of Jr. Novice Mid students was 3.43, indicating that a number of these students rated themselves higher than did the one student in the Jr. Novice High range.

Table 5: SOPA Listening Comprehension Levels and Student Self-Assessment Mean Totals for Spanish

SOPA Listening Comprehension Levels	Number of Students Per Level	Student-Self Assessment	
		Mean	Standard Deviation
Jr. Novice Low	5	22.20	2.16
Jr. Novice Mid	28	23.07	2.99
Jr. Novice High	26	23.38	3.76

Table 6: SOPA Fluency Levels and Student Self-Assessment Mean Totals for Spanish

SOPA Fluency Levels	Number of Students Per Level	Student-Self Assessment	
		Mean	Standard Deviation
Jr. Novice Low	24	22.04	2.82
Jr. Novice Mid	34	23.91	3.43
Jr. Novice High	1	23.00	0.00

The correlation between SOPA comprehension ratings and SSA total scores was .12 (NS, $p = .361$) for Spanish. The correlation between SOPA oral fluency ratings and SSA total scores was .29 (sig., $p = .025$).

While there was little or no correlation between the Spanish SOPA ratings and SSA total scores, it is clear that there was also not a wide range of scores (as there was for example, in the Chinese data). The relative weakness of these correlations is a function of the homogenous nature of the Spanish data set. Even with this relatively homogeneous group, however, we see that the comparisons of SOPA ratings and corresponding mean total SSA scores provide some support to the hypothesis that the higher the SOPA rating (listening or oral fluency) is, the higher the total score on the SSA.

SOPA Levels and Teacher's FLOSEM Ratings

Table 7 illustrates, without exception, that the higher the student SOPA listening comprehension level for Spanish was, the higher the FLOSEM comprehension. SOPA listening comprehension levels ranged from Jr. Novice Low to Jr. Novice High. Corresponding FLOSEM comprehension mean ratings climbed steadily from 2.00 to 3.00.

Table 7: SOPA Listening Comprehension Levels and FLOSEM Comprehension Mean Ratings for Spanish

SOPA Listening Comprehension Levels	Number of Students	FLOSEM Comprehension
Jr. Novice Low	5	2.00 (0.00)
Jr. Novice Mid	29	2.62 (0.49)
Jr. Novice High	26	3.00 (0.28)

Table 8 illustrates that the higher the SOPA oral fluency level was, the higher the FLOSEM fluency level. SOPA oral fluency levels ranged from Jr. Novice Low to Junior Novice High. FLOSEM fluency ratings ranged from 1.54 to 2.00. Comparisons between SOPA oral fluency levels and FLOSEM vocabulary, pronunciation, grammar, and total fluency show the same trend.

Table 8: SOPA Fluency Levels and FLOSEM Fluency, Vocabulary, Pronunciation, Grammar, and Fluency Total Means for Spanish

SOPA Fluency Levels	Number of Students	FLOSEM Fluency	FLOSEM Vocabulary	FLOSEM Pronunciation	FLOSEM Grammar	FLOSEM Fluency Total
Jr. Novice Low	24	1.54 (0.51)	1.50 (0.51)	2.71 (0.55)	1.63 (0.49)	9.79 (2.23)
Jr. Novice Mid	35	1.94 (0.34)	1.97 (0.30)	3.23 (0.55)	2.03 (0.30)	12.11 (1.51)
Jr. Novice High	1	2.00 (0.00)	2.00 (0.00)	4.00 (0.00)	2.00 (0.00)	13.00 (0.00)

The correlation between SOPA listening comprehension levels and FLOSEM comprehension was .57 (sig., p = .000) for Spanish. The correlation between Spanish SOPA oral fluency levels and FLOSEM fluency ratings was .44 (sig., p = .000). Correlations between Spanish SOPA oral fluency levels and other FLOSEM components (all considered part of fluency) were as follows: vocabulary .51 (sig., p = .000), pronunciation .46 (sig., p = .000), and grammar .46 (sig., p = .000).

As in the Chinese study, the correlations between SOPA levels and FLOSEM ratings and comparisons of SOPA levels and FLOSEM mean ratings provide support for the validity of the SOPA as an oral assessment instrument. The hypothesis that the higher the SOPA listening comprehension or oral fluency rating, the higher the corresponding (composite or individual) FLOSEM rating is supported.

Subset C: French

SOPA Ratings and Student Self-Assessment Scores

For French, SOPA listening comprehension ratings ranged from Jr. Novice Low to Jr. Intermediate Low and SSA totals ranged from 19.46 to 24.70 (see Table 9). SOPA oral fluency ratings ranged from Jr. Novice Low to Jr. Intermediate Low, and SSA totals ranged from 22.08 to 25.50 (see Table 10).

Table 9: SOPA Listening Comprehension Levels and Student Self-Assessment Mean Totals for French

SOPA Listening Comprehension Levels	Number of Students Per Level	Student-Self Assessment	
		Mean	Standard Deviation
Jr. Novice Low	13	19.46	2.78
Jr. Novice Mid	39	23.15	2.45
Jr. Novice High	46	23.37	2.35
Jr. Intermediate Low	10	24.70	2.05

Table 10: SOPA Fluency Levels and Student Self-Assessment Mean Totals for French

SOPA Fluency Levels	Number of Students Per Level	Student-Self Assessment	
		Mean	Standard Deviation
Jr. Novice Low	37	22.08	3.08
Jr. Novice Mid	46	23.17	2.70
Jr. Novice High	23	23.65	1.92
Jr. Intermediate Low	2	25.50	0.71

The correlation between the SOPA listening comprehension ratings and SSA total scores was .33 (sig., p = .001) for French. The correlation between the SOPA oral fluency ratings and SSA was .21 (sig., p = .033).

While correlations were low, both the correlations and the increasing trend in mean SSA ratings provide some support to the hypothesis that the higher the SOPA rating is (listening comprehension or oral fluency), the higher the total score on the SSA. While the SOPA levels Jr. Novice Low and Jr. Intermediate Low were clearly differentiated on the SSA, the two middle categories for this group were not.

SOPA Levels and Teachers' FLOSEM Ratings

In general, all French teachers rated students higher on the FLOSEM who had received higher ratings on the SOPA. However, due to individual differences among the French teachers in how they interpreted the levels of the FLOSEM, the FLOSEM ratings could not be combined across teachers and programs. Therefore, comparisons of mean results for French are presented by teacher.

French Teacher A

As Table 11 illustrates, French Teacher A assigned the following average FLOSEM comprehension ratings to her students: for students at the SOPA Jr. Novice Low level, 2.00; for Jr. Novice Mid students, 1.60; and for Jr. Novice High students, 2.69. The standard deviations for FLOSEM comprehension ratings for the Jr. Novice Low and Jr. Novice Mid students (.50 and .53 respectively) indicate that individual scores varied from this pattern.

Table 11: SOPA Comprehension Levels and FLOSEM Comprehension Means (French Teacher A)

SOPA Listening Comprehension Levels	Number of Students	FLOSEM Comprehension
Jr. Novice Low	3	2.00 (0.50)
Jr. Novice Mid	14	1.60 (0.53)
Jr. Novice High	13	2.69 (0.48)

Table 12 shows that for SOPA oral fluency, the higher the SOPA rating was, the higher the FLOSEM fluency rating. SOPA oral fluency ratings ranged from Jr. Novice Low to Jr. Novice High. Corresponding FLOSEM fluency levels ranged from 1.50 to 2.17. For FLOSEM fluency totals and FLOSEM fluency components, student mean ratings showed similar trends with the exception of grammar, which stayed constant.

Table 12: SOPA Fluency Levels and FLOSEM Fluency, Vocabulary, Pronunciation, Grammar, and Fluency Total Means (French Teacher A)

SOPA Fluency Levels	Number of Students	FLOSEM Fluency	FLOSEM Vocabulary	FLOSEM Pronunciation	FLOSEM Grammar	FLOSEM Fluency Total
Jr. Novice Low	15	1.50 (0.53)	1.00 (0.00)	1.80 (0.41)	1.00 (0.00)	7.17 (1.41)
Jr. Novice Mid	12	1.88 (0.61)	1.12 (0.23)	2.13 (0.57)	1.00 (0.00)	8.33 (2.06)
Jr. Novice High	3	2.17 (0.58)	1.33 (0.29)	2.50 (0.50)	1.00 (0.00)	10.00 (1.80)

French Teacher B

As Table 13 shows, while French Teacher B's students were assigned SOPA listening comprehension levels of Jr. Novice Low to Jr. Novice High, actual mean FLOSEM comprehension scores are the same (2.43) for students at both the Jr. Novice Mid and Jr. Novice High level. On the FLOSEM, Teacher B did not distinguish these two groups in comprehension as the SOPA had.

Table 13: SOPA Comprehension Levels and FLOSEM Comprehension Means (French Teacher B)

SOPA Listening Comprehension Levels	Number of Students	FLOSEM Comprehension
Jr. Novice Low	9	1.89 (0.33)
Jr. Novice Mid	14	2.43 (0.76)
Jr. Novice High	7	2.43 (0.53)

Table 14 shows that as SOPA oral fluency levels are higher, so are FLOSEM fluency ratings for Teacher B. SOPA oral fluency levels ranged from Jr. Novice Low to Jr. Novice Mid. FLOSEM fluency levels ranged from 1.11 to 1.55. FLOSEM fluency total mean ratings and FLOSEM fluency components followed similar trends. Thus, regarding fluency, Teacher B distinguished between the students' relative abilities in the same way as the SOPA.

Table 14: SOPA Fluency Levels and FLOSEM Fluency, Vocabulary, Pronunciation, Grammar, and Fluency Total Means for French (French Teacher B)

SOPA Fluency Levels	Number of Students	FLOSEM Fluency	FLOSEM Vocabulary	FLOSEM Pronunciation	FLOSEM Grammar	FLOSEM Fluency Total
Jr. Novice Low	19	1.11 (0.32)	1.68 (0.67)	2.79 (0.71)	1.89 (0.46)	9.53 (2.27)
Jr. Novice Mid	11	1.55 (0.52)	2.09 (0.94)	3.36 (0.81)	2.45 (0.52)	12.09 (3.08)

French Teacher C

Table 15 shows that French Teacher C's students were assigned listening comprehension levels of Jr. Novice Mid to Jr. Intermediate Low on the SOPA. Corresponding FLOSEM comprehension scores range from 1.93 to 2.38.

Table 15: SOPA Comprehension Levels and FLOSEM Comprehension Means (French Teacher C)

SOPA Listening Comprehension Levels	Number of Students	FLOSEM Comprehension
Jr. Novice Mid	10	1.93 (0.33)
Jr. Novice High	14	2.00 (0.17)
Jr. Intermediate Low	6	2.38 (0.21)

Table 16 shows a similar trend for oral fluency levels, FLOSEM fluency, fluency components, and total scores. SOPA oral fluency levels ranged from Jr. Novice Low to Jr. Intermediate Low. FLOSEM fluency scores climbed from 1.00 to 1.50. FLOSEM fluency totals and FLOSEM pronunciation ratings assigned by Teacher C showed a very similar trend while ratings for vocabulary and grammar showed little or no variation.

Table 16: SOPA Fluency Levels and FLOSEM Fluency, Vocabulary, Pronunciation, Grammar, and Fluency Total Means (French Teacher C)

SOPA Fluency Levels	Number of Students	FLOSEM Fluency	FLOSEM Vocabulary	FLOSEM Pronunciation	FLOSEM Grammar	FLOSEM Fluency Total
Jr. Novice Low	1	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	6.00 (0.00)
Jr. Novice Mid	11	1.27 (0.34)	1.05 (0.15)	1.09 (0.20)	1.00 (0.00)	6.34 (0.73)
Jr. Novice High	16	1.52 (0.39)	1.03 (0.13)	1.22 (0.26)	1.00 (0.00)	6.84 (0.86)
Jr. Intermediate Low	2	1.50 (0.00)	1.00 (0.00)	1.50 (0.00)	1.00 (0.00)	7.50 (0.00)

French Teacher D

As Table 17 shows, French Teacher D's students were assigned SOPA listening comprehension levels that ranged from Jr. Novice Low to Jr. Intermediate Low. Corresponding FLOSEM comprehension ratings climb from 1.33 to 3.00.

Table 17: SOPA Comprehension Levels and FLOSEM Comprehension Means (French Teacher D)

SOPA Listening Comprehension Levels	Number of Students	FLOSEM Comprehension
Jr. Novice Low	1	1.00 (0.00)
Jr. Novice Mid	3	1.33 (0.58)
Jr. Novice High	16	2.19 (0.40)
Jr. Intermediate Low	4	3.00 (0.00)

Table 18 shows a similar relationship between SOPA fluency levels and corresponding FLOSEM fluency, pronunciation, grammar, vocabulary, and fluency total ratings. The higher the SOPA fluency level, the higher the FLOSEM rating.

Table 18: SOPA Fluency Levels and FLOSEM Fluency, Vocabulary, Pronunciation, Grammar, and Fluency Total Means (French Teacher D)

SOPA Fluency Levels	Number of Students	FLOSEM Fluency	FLOSEM Vocabulary	FLOSEM Pronunciation	FLOSEM Grammar	FLOSEM Fluency Total
Jr. Novice Low	2	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	5.00 (0.00)
Jr. Novice Mid	14	2.00 (0.39)	1.93 (0.27)	2.79 (0.43)	2.00 (0.39)	10.71 (1.49)
Jr. Novice High	8	2.88 (0.35)	2.50 (0.53)	3.50 (0.53)	2.50 (0.53)	14.13 (2.10)

The correlation between SOPA listening comprehension levels and FLOSEM comprehension ratings was .41 (sig., p = .000) for all French teachers combined. The correlation between SOPA oral fluency levels and FLOSEM fluency ratings was .42 (sig., p = .000). Correlations between French SOPA oral fluency levels and other FLOSEM components (all considered part of fluency) were as follows: vocabulary .066 (NS, p = .48), pronunciation -.122 (NS, p = .194), and grammar -.060 (NS, p = .529).

As in the Spanish study, for most of the French classes there was little variation between proficiency levels assigned by the SOPA. Nevertheless, the data shows support, particularly for the FLOSEM, that the SOPA is assessing oral language skills, because students with higher SOPA levels received higher ratings from their teachers.

Summary of Results

SOPA Levels and Student Self-Assessment Scores

For all languages, the higher the SOPA level was, the higher the SSA mean total scores. Correlations between SOPA ratings and SSA scores were most compelling for Chinese, followed by French, and then Spanish. Correlations between the SOPA listening comprehension levels and SSA totals were strong for Chinese and French but less so for Spanish. Similarly, correlations between the SOPA fluency levels and SSA were strong for Chinese and French but weaker for Spanish.

SOPA Levels and FLOSEM Ratings

Correlations between SOPA ratings and FLOSEM ratings were most compelling for Chinese, followed by Spanish, and then French. For SOPA listening comprehension levels and FLOSEM comprehension ratings, the correlations were strong for Chinese and Spanish and weak for French. For SOPA fluency and FLOSEM fluency ratings, the correlations were strongest for Chinese and Spanish, followed by French.

Conclusion

The interpretation of the FLES results (correlations and comparison of means) provides evidence for the validity of the FLES SOPA as an assessment of oral proficiency and listening comprehension through the relationships demonstrated between SOPA outcomes, student self-assessment of oral language skills, and teacher ratings of oral language skills.

VI. Methodology – Study 2: Immersion SOPA Validation

Subjects

The immersion SOPA validation study required the administration of the immersion SOPA to representative immersion programs in elementary schools in the United States. A first site (F), representing the total immersion model, was identified during a week-long Performance Assessment Institute at the Iowa State University National K-12 Foreign Language Assessment Institute in which one of its administrators participated. A second site (G), representing the two-way partial immersion model, was selected because its staff had previously trained in using the SOPA as part of their own program evaluation efforts.

Total Immersion – Site F

This site is a total immersion program in the mid-Atlantic region. Students are selected by lottery to participate in the program. Instruction is entirely in French until fifth grade. In fifth grade, students study in English 20% of the time. The population tested consisted of third grade students who were from English-speaking homes (n = 22), or homes where French (n = 4) or French and Creole (n = 6) were spoken. Students were randomly selected from the two third grade classes. One teacher is a native speaker of French and the other is fluent in French.

Two-Way Partial Immersion – Site G

Students in this program receive 50% of instruction in Spanish and 50% in English in Grades K-5. In this program, located in the mid-Atlantic region, the student body is comprised of approximately half native English speakers and half native Spanish speakers. Twenty-five native Spanish speakers and 25 native English speakers in fourth grade were randomly selected to participate in this study. All three teachers are fluent in Spanish and English. Two of the teachers are native speakers of Spanish.

Chart B: Language Programs Participating in SOPA Validation Study
IMMERSION PROGRAMS

Site Location	Native Speakers of Target Languages (students)	Program Type	Student Selection	Grade	# of Students per Grade	Teachers' Language	Special Characteristics
Spanish							
Mid–Atlantic	50% Spanish	Two–Way Partial Immersion K-5 (50% in Spanish)	Full range (random)	4	50	Native (2) Non-native/fluent (1)	Half of the students speak or are exposed to Spanish at home.
French							
Mid–Atlantic	12.5% French 18.75% French/Creole	Total Immersion K-4 (100% in French) Grade 5 (80% in French)	Full range (random)	3	32	Native (1) Non-native/fluent (1)	Almost one-third of the students tested have French (4) or French/Creole (6) language exposure at home.

Instrumentation

The immersion validation study plan was identical to the less intensive instruction plan. Students were interviewed using the immersion SOPA, rated by their teachers using the FLOSEM, and rated themselves using an immersion version of the Student Self-Assessment (SSA). This SSA was developed with direct input from immersion teachers at both sites. Two versions of the SSA were available, one in English and one in the target language. Both versions asked students to rate their ability in English and in the target language by indicating to what extent they were able to successfully understand or communicate in a broad range of situations. Finally, background information about each program was collected through the completion of a questionnaire at each site. It was anticipated that such information would aid the interpretation of assessment results.

Procedures

Immersion teachers at each site were given additional training in the use of the latest version of the SOPA. These teachers then interviewed a randomly-selected group of students at each site, aided CAL staff in the collection of background information, and facilitated the administration of the SSA and the FLOSEM. CAL staff who had been involved in the development of the immersion SOPA served as raters at the sites.

Data Analysis

Feedback to the Sites

For each site, student averages for all instruments were calculated, discussed, and interpreted in a written report to the school. Statistically significant differences in student performance due to language background were reported. These reports are not included here since they report results according to site rather than according to language and program type.

Immersion Validation Study

For purposes of immersion SOPA validation, the data were handled differently. The emphasis in the validation study was not on reporting average student performance on each of the instruments, but on looking at the relationship between student performance on the SOPA, the SSA and the FLOSEM and seeing if this information confirms the validity of the SOPA. Due to differences in type of immersion program, language studied, and students' home language, the data for each site was analyzed separately.

For the SSA, correlational studies were conducted to address the following hypotheses:

1. The higher the SSA in the target language, the higher the SOPA comprehension score should be.

2. The higher the SSA in the target language, the higher the SOPA fluency score should be.

To test the first hypothesis, SSA target language scores were correlated with SOPA comprehension scores. To confirm the second hypothesis, SSA target language scores were correlated with SOPA fluency scores. The correlation between SSA total scores and SOPA fluency was expected to be higher than SSA total scores and SOPA comprehension since most of the SSA items relate to fluency.

For the FLOSEM, the analysis involves not only oral fluency and listening comprehension but also grammar and vocabulary for the immersion program data. The FLOSEM rates pronunciation as a separate variable. The SOPA considers pronunciation to be part of fluency. In addition, it was necessary to note that the immersion program SOPA uses a 9-level rating matrix (junior novice low to junior superior) while the FLOSEM uses a 6-level rating matrix (1 to 6).

The following hypotheses were examined:

1. The higher the FLOSEM comprehension score, the higher the SOPA comprehension score should be.
2. The higher the FLOSEM fluency score, the higher the SOPA oral fluency score should be.
3. The higher the FLOSEM grammar score, the higher the SOPA grammar score should be.
4. The higher the FLOSEM vocabulary score, the higher the SOPA vocabulary score should be.
5. The higher the FLOSEM pronunciation score, the higher the SOPA fluency score should be.

To confirm the first hypothesis, FLOSEM comprehension and SOPA comprehension were correlated with the assumption that this correlation should be higher than the correlation between FLOSEM comprehension and SOPA fluency or other skill areas (grammar and vocabulary). For the second, FLOSEM fluency and SOPA fluency were correlated with the assumption that this correlation should be higher than the correlation between FLOSEM fluency and SOPA comprehension or other skill areas. For the third, FLOSEM grammar and SOPA grammar scores were correlated with the assumption that this correlation should be higher than the correlation between FLOSEM grammar and SOPA comprehension, fluency, or vocabulary. For the fourth, FLOSEM vocabulary and SOPA vocabulary scores were correlated with the assumption that this correlation should be higher than the correlation between FLOSEM vocabulary and SOPA comprehension, fluency, or grammar. For the fifth, FLOSEM pronunciation and SOPA fluency scores were correlated with the assumption that this correlation should be higher

than the correlation between FLOSEM pronunciation and SOPA comprehension, grammar, or vocabulary.

Unlike the SOPA, which was rated by CAL staff or consultants who had received intensive training together, the FLOSEM was rated by each classroom teacher at each site. These teachers were trained separately by different CAL consultants. Therefore, greater variation in the interpretation and application of the FLOSEM observation matrix was expected than in the assignment of SOPA ratings.

Background Variables Considered

Average scores on all instruments and correlations between scores were examined according to the home language background of students (English or home exposure to other languages). In addition, student perceptions of language ability in their native and second language versus SOPA scores were examined through the administration of an SSA, which asked students to rate their ability in both English and the target language.

VII. Results – Study 2: Immersion SOPA Validation

Results of data analysis are presented by site: French total immersion program (Site F) and Spanish two-way immersion program (Site G). Correlations between the SOPA ratings and the Student Self-Assessment (SSA) total scores, the SOPA ratings (comprehension, fluency, vocabulary, and grammar) and the FLOSEM ratings (comprehension, fluency, vocabulary, pronunciation, and grammar), along with a comparison of mean scores (and their standard deviations) on all instruments are presented and discussed separately for the two data subsets.

Subset F: Total Immersion

SOPA Ratings and Student Self-Assessment Scores

For Subset F, SOPA listening comprehension ratings ranged from Jr. Intermediate Mid to Jr. Advanced High and SSA totals from 40.00 to 42.09 (see Table 19).

Table 19: Comparison SOPA Listening Comprehension Levels and Student Self-Assessment Totals for Total Immersion

SOPA Listening Comprehension Level	Number of Students	Mean SSA French Total	Standard Deviation
Jr. Intermediate Mid	1	40.00	0.00
Jr. Intermediate High	1	42.00	0.00
Jr. Advanced	26	42.09	3.16
Jr. Advanced High	2	40.50	2.12

SOPA fluency ratings were within the same range as for listening comprehension, while SSA totals ranged from 39.00 to 42.79 (see Table 20).

Table 20: Comparison of SOPA Fluency Levels and Student Self-Assessment Totals for Total Immersion

SOPA Fluency Level	Number of Students	Mean SSA French Total	Standard Deviation
Jr. Intermediate Mid	5	42.79	1.91
Jr. Intermediate High	6	40.56	3.01
Jr. Advanced	18	42.28	3.23
Jr. Advanced High	1	39.00	0.00

SOPA grammar ratings ranged from Jr. Intermediate High to Jr. Advanced High and SSA totals from 40.90 to 43.00 (see Table 21).

Table 21: Comparison of SOPA Grammar Levels and Student Self-Assessment Totals for Total Immersion

SOPA Grammar Level	Number of Students	Mean SSA French Total	Standard Deviation
Jr. Intermediate High	11	40.90	3.26
Jr. Advanced	17	42.43	2.92
Jr. Advanced High	2	43.00	1.41

SOPA vocabulary ratings ranged from Jr. Intermediate Mid to Jr. Advanced. SSA totals ranged from 40.64 to 42.92 (see Table 22).

Table 22: Comparison SOPA Vocabulary Levels and Student Self-Assessment Totals for Total Immersion

SOPA Vocabulary Level	Number of Students	Mean SSA French Total	Standard Deviation
Jr. Intermediate Mid	8	41.99	2.38
Jr. Intermediate High	10	40.64	4.00
Jr. Advanced	12	42.92	2.15

The correlations between SOPA skill levels and SSA total scores were not significant. The correlations were as follows: -.025 (NS, p = .896) for SOPA comprehension, .085 (NS, p = .655) for SOPA fluency; .25 (NS, p = .182) for SOPA grammar; and .20 (NS, p = .285) for SOPA vocabulary.

The correlation between SOPA grammar ratings and SSA total scores was the strongest (.25), but was not significant. A comparison of means between the two instruments also showed some support for the hypothesis that as SOPA grammar scores increase, so do SSA total scores. For all other comparisons, SSA total scores did not steadily increase as SOPA ratings increased.

SOPA Levels and Teachers' FLOSEM Ratings

Table 23 shows the SOPA listening comprehension levels achieved, the number of students who achieved each level, their mean FLOSEM comprehension rating, and its standard deviation. As Table 23 illustrates, the higher the SOPA listening comprehension rating, the higher the FLOSEM comprehension ratings for French. For comprehension on the FLOSEM, the teachers' ratings did not distinguish between Jr. Intermediate Mid and Jr. Intermediate High as the SOPA had.

Table 23: Comparison of SOPA Listening Comprehension Levels and FLOSEM Listening Comprehension Ratings for Total Immersion

SOPA Listening Comprehension Level	Number of Students	FLOSEM Mean Ratings	Standard Deviation
Jr. Intermediate Mid	1	3.00	0.00
Jr. Intermediate High	1	3.00	0.00
Jr. Advanced	28	4.68	1.06
Jr. Advanced High	2	5.50	0.71

Table 24 shows that as SOPA fluency levels increased, FLOSEM ratings generally increased. The teachers' FLOSEM ratings were slightly lower for Jr. Intermediate High students.

Table 24: Comparison of SOPA Fluency Levels and FLOSEM Fluency Ratings for Total Immersion

SOPA Fluency Level	Number of Students	FLOSEM Mean Ratings	Standard Deviation
Jr. Intermediate Mid	6	3.67	1.21
Jr. Intermediate High	6	3.50	1.22
Jr. Advanced	18	4.22	0.94
Jr. Advanced High	2	5.00	0.00

The SOPA includes pronunciation as one of the characteristics of fluency. Table 25 shows that as SOPA fluency levels increased, FLOSEM pronunciation ratings generally increased. Teachers' ratings for FLOSEM pronunciation did not distinguish between students at the Jr. Advanced and Jr. Advanced High levels.

Table 25: Comparison of SOPA Fluency Levels and FLOSEM Pronunciation Ratings for Total Immersion

SOPA Fluency Level	Number of Students	FLOSEM Pronunciation Mean Ratings	Standard Deviation
Jr. Intermediate Mid	6	3.50	0.55
Jr. Intermediate High	6	3.83	0.41
Jr. Advanced	18	4.00	0.48
Jr. Advanced High	2	4.00	0.00

A comparison of SOPA grammar levels and FLOSEM grammar ratings show that teachers' FLOSEM grammar ratings for Jr. Advanced students were slightly higher than for Jr. Advanced High students (see Table 26).

Table 26: Comparison of SOPA Grammar Levels and FLOSEM Grammar Ratings for Total Immersion

SOPA Grammar Level	Number of Students	FLOSEM Mean Ratings	Standard Deviation
Jr. Intermediate High	12	3.00	0.85
Jr. Advanced	17	3.82	0.64
Jr. Advanced High	3	3.67	0.58

Finally, comparing SOPA vocabulary levels and FLOSEM vocabulary ratings reveals that as SOPA levels increased, so did FLOSEM ratings (see Table 27).

Table 27: Comparison of SOPA Vocabulary Levels and FLOSEM Vocabulary Ratings for Total Immersion

SOPA Vocabulary Level	Number of Students	FLOSEM Mean Ratings	Standard Deviation
Jr. Intermediate Mid	9	3.55	0.88
Jr. Intermediate High	10	3.90	0.88
Jr. Advanced	13	4.38	0.77

The correlations between SOPA levels and FLOSEM ratings are listed in Table 28. It was expected that correlations between parallel skills would be significant. Indeed, correlations between SOPA and FLOSEM listening comprehension and grammar ratings were all significant. The correlations between SOPA and FLOSEM fluency ratings and SOPA and FLOSEM vocabulary ratings were not significant for this data.

In addition, correlations between FLOSEM pronunciation ratings and the SOPA fluency levels were expected. Positive, significant correlations were found between FLOSEM pronunciation ratings and SOPA fluency, listening comprehension, and grammar levels. The correlation between SOPA vocabulary levels and FLOSEM pronunciation ratings was not significant.

Significant correlations were also found between differing skills assessed by the two instruments. SOPA fluency correlated significantly with FLOSEM listening, vocabulary, and grammar. Significant correlations were found between SOPA listening comprehension and FLOSEM vocabulary and grammar. SOPA vocabulary correlated significantly with FLOSEM grammar.

Table 28: Correlations Between SOPA Levels and FLOSEM Ratings for Total Immersion

SOPA Skills	FLOSEM Fluency	FLOSEM Listening	FLOSEM Vocabulary	FLOSEM Pronunciation	FLOSEM Grammar
SOPA Fluency	.33 (NS, p = .064)	.43 (sig., p = .012)	.52 (sig., p = 002)	.36 (sig., p = .042)	.45 (sig., p = .009)
SOPA Listening Comprehension	.18 (NS, p = .321)	.40 (sig., p = .023)	.39 (sig., p.= .027)	.38 (sig., p = .032)	.34 (NS, p = .056)
SOPA Vocabulary	.08 (NS, p = .101)	.29 (NS, p = .655)	.39 (sig., p = .027)	.18 (NS, p = .308)	.38 (sig., p = .032)
SOPA Grammar	.31 (NS, p = .080)	.46 (sig., p = .008)	.48 (sig., p =.005)	.43 (sig., p = .013)	.41 (sig., p = .019)

Correlations and comparisons of means between SOPA levels and FLOSEM ratings generally support the hypothesis that the two instruments are assessing the same thing since for most students, the higher the SOPA rating, the higher the corresponding FLOSEM rating. Differences in the degree of experience with the FLOSEM may account for variation in the way students were rated by their teachers. An experienced CAL SOPA rater rated all SOPA interviews, whereas the FLOSEM ratings were assigned by two teachers who had limited training and no prior experience rating students with the scale.

Subset G: Two-Way Partial Immersion

SOPA Ratings and Student Self-Assessment Scores

For Subset G, Table 29 shows that SOPA listening comprehension ratings ranged from Junior Intermediate Low to Junior Advanced. SSA scores, with the exception of one level, increased as SOPA levels increased.

Table 29: Comparison of SOPA Listening Comprehension Levels and Student Self-Assessment Totals for Two-Way Immersion

SOPA Listening Comprehension Level	Number of Students	Mean SSA Spanish Total	Standard Deviation
Jr. Intermediate Low	1	42.86*	0.00
Jr. Intermediate Mid	4	36.25	3.10
Jr. Intermediate High	4	40.93	3.56
Jr. Advanced	39	41.68	2.31

*This is an estimated score. The student's response to one SSA question was missing. The average of the rest of the student's SSA responses was used to estimate this missing value.

As Table 30 illustrates, SOPA fluency levels ranged from Junior Novice Mid to Junior Advanced High. SSA mean totals did not consistently increase as SOPA fluency levels increased. Standard deviations between SSA total scores

ranged from 1.71 to 4.92, suggesting that there was a fair amount of variability in how individuals rated themselves on the SSA.

Table 30: Comparison of SOPA Fluency Levels and Student Self-Assessment Totals for Two-Way Immersion

SOPA Fluency Level	Number of Students	Mean SSA Spanish Total	Standard Deviation
Jr. Novice Mid	1	42.86*	0.00
Jr. Novice High	1	36.00	0.00
Jr. Intermediate Low	4	38.25	4.92
Jr. Intermediate Mid	6	41.25	2.80
Jr. Intermediate High	11	41.14	2.24
Jr. Advanced	21	42.30	2.27
Jr. Advanced High	4	39.25	1.71

*This is an estimated score. The student's response to one SSA question was missing. The average of the rest of the student's SSA responses was used to estimate this missing value.

Table 31 shows that SSA mean total scores did not consistently increase as SOPA grammar levels increased. In addition, standard deviations show that there was a good amount of variability in how students rated themselves on the SSA (SD. = 1.61 to 5.65).

Table 31: Comparison of SOPA Grammar Levels and Student Self-Assessment Totals for Two-Way Immersion

SOPA Grammar Level	Number of Students	Mean SSA Spanish Total	Standard Deviation
Jr. Novice Mid	1	42.86*	0.00
Jr. Novice High	2	40.00	5.66
Jr. Intermediate Low	4	37.75	4.19
Jr. Intermediate Mid	9	41.60	2.91
Jr. Intermediate High	8	40.70	1.61
Jr. Advanced	13	42.72	2.07
Jr. Advanced High	11	42.73	2.61

*This is an estimated score. The student's response to one SSA question was missing. The average of the rest of the student's SSA responses was used to estimate this missing value.

A comparison of SOPA vocabulary levels and SSA totals reveals, with the exception of two levels, that as SOPA vocabulary levels increased so did SSA totals (see Table 32). Standard deviations suggest, however, quite a bit of variation in the way students rated themselves (SD. = 1.20 to 6.11).

Table 32: Comparison of SOPA Vocabulary Levels and Student Self-Assessment Totals for Two-Way Immersion

SOPA Vocabulary Level	Number of Students	Mean SSA Spanish Total	Standard Deviation
Jr. Novice Mid	1	42.86*	0.00
Jr. Novice High	3	37.33	6.11
Jr. Intermediate Low	6	40.12	3.05
Jr. Intermediate Mid	6	42.79	1.20
Jr. Intermediate High	12	41.27	2.33
Jr. Advanced	20	41.48	2.51

*This is an estimated score. The student's response to one SSA question was missing. The average of the rest of the student's SSA responses was used to estimate this missing value.

Overall, the comparison of trends in SOPA skill levels assigned and student SSA scores in Subset G does not offer any conclusive findings, since as SOPA skill levels increase, SSA total scores increase at some levels and decrease at others.

In addition, the correlations between SOPA skill levels and SSA total scores were not significant. The correlations were as follows: .279 (NS, p = .055) for SOPA comprehension; .145 (NS, p = .323) for SOPA fluency; .106 (NS, p = .471) for SOPA grammar; and .087 (NS, p = .556) for SOPA vocabulary. Thus, correlations between the SOPA skill levels and SSA total scores for Subset G did not support the hypothesis that the higher the SOPA rating, the higher the total score on the SSA.

SOPA Levels and Teachers' FLOSEM Ratings

Table 33 shows the SOPA listening comprehension levels achieved, the number of students who achieved each level, their mean FLOSEM comprehension rating, and its standard deviation. As Table 33 illustrates, the higher the SOPA listening comprehension level, the higher the FLOSEM listening comprehension rating.

Table 33: Comparison of SOPA Comprehension Levels and FLOSEM Comprehension Ratings for Two-Way Immersion

SOPA Listening Comprehension Level	Number of Students	FLOSEM Mean Ratings	Standard Deviation
Jr. Intermediate Low	1	2.00	0.00
Jr. Intermediate Mid	5	3.20	1.10
Jr. Intermediate High	4	4.75	0.96
Jr. Advanced	40	5.77	0.53

Table 34 shows a generally similar trend between SOPA levels and FLOSEM ratings for fluency, with two exceptions. First, teacher FLOSEM ratings do not distinguish between Jr. Novice Mid and Jr. Novice High for

fluency. Second, teacher FLOSEM ratings were also slightly higher for students at the SOPA Jr. Intermediate Mid than at the Jr. Intermediate High level. Standard deviations for these ratings indicate a fair amount of variation in student FLOSEM ratings.

Table 34: Comparison of SOPA Fluency Levels and FLOSEM Fluency Ratings for Two-Way Immersion

SOPA Fluency Level	Number of Students	FLOSEM Mean Ratings	Standard Deviation
Jr. Novice Mid	1	2.00	0.00
Jr. Novice High	1	2.00	0.00
Jr. Intermediate Low	5	3.00	0.71
Jr. Intermediate Mid	6	4.00	0.63
Jr. Intermediate High	11	3.82	0.60
Jr. Advanced	22	5.77	0.68
Jr. Advanced High	4	6.00	0.00

A comparison of SOPA grammar levels and FLOSEM grammar ratings shows that on the FLOSEM, teacher ratings were slightly lower for students at the Jr. Intermediate High level than at the Jr. Intermediate Mid level. However, a look at the standard deviations for these ratings (see Table 35) shows that there was variability in the ratings assigned for students at this level.

Table 35: Comparison of SOPA Grammar Levels and FLOSEM Grammar Ratings for Two-Way Immersion

SOPA Grammar Level	Number of Students	FLOSEM Mean Ratings	Standard Deviation
Jr. Novice Mid	1	2.00	0.00
Jr. Novice High	3	2.67	0.58
Jr. Intermediate Low	4	2.75	0.50
Jr. Intermediate Mid	9	3.67	0.50
Jr. Intermediate High	8	3.50	0.76
Jr. Advanced	14	5.21	0.58
Jr. Advanced High	11	5.45	0.52

Table 36 shows that as SOPA vocabulary levels increased so did FLOSEM vocabulary ratings.

Table 36: Comparison of SOPA Vocabulary Levels and FLOSEM Vocabulary Ratings for Two-Way Immersion

SOPA Vocabulary Level	Number of Students	FLOSEM Mean Ratings	Standard Deviation
Jr. Novice Mid	1	2.00	0.00
Jr. Novice High	3	2.66	0.58
Jr. Intermediate Low	7	3.57	0.50
Jr. Intermediate Mid	6	4.00	0.00
Jr. Intermediate High	12	4.25	1.05
Jr. Advanced	21	5.76	0.43

Overall, comparison of SOPA skill levels and corresponding FLOSEM ratings strongly support the hypothesis that the two instruments are measuring the same thing, since the higher the SOPA level, the higher the FLOSEM rating assigned.

The correlations between SOPA levels and FLOSEM ratings are listed below in Table 37. It was expected that correlations between parallel skills would be significant. In fact, correlations between SOPA levels and corresponding FLOSEM ratings were high and significant for all skills.

In addition, it was expected that FLOSEM pronunciation ratings would correlate significantly with SOPA fluency levels. Results show that the FLOSEM pronunciation ratings for the Spanish two-way immersion site correlate highly with grammar, fluency, and vocabulary.

Similar to the total immersion data, significant correlations were found between the ratings for different skills. In the two-way immersion data, all correlations between like as well as different ratings for SOPA and FLOSEM skills were significant ($p = .000$).

Table 37: Correlations Between SOPA Levels and FLOSEM Ratings for Two-Way Immersion

SOPA Skill	FLOSEM Fluency	FLOSEM Listening	FLOSEM Vocabulary	FLOSEM Pronunciation	FLOSEM Grammar
SOPA Fluency	.87	.67	.78	.83	.84
SOPA Listening Comprehension	.62	.72	.62	.55	.58
SOPA Vocabulary	.85	.66	.84	.79	.78
SOPA Grammar	.88	.67	.79	.84	.85

Thus, for the two-way immersion data, high and significant correlations between SOPA levels and FLOSEM ratings were found. This outcome provides strong evidence for the validity of the SOPA as an assessment of oral language skills. The hypothesis was upheld that the higher the SOPA skill level, the higher the corresponding FLOSEM rating.

Summary of Results

SOPA Levels and Student Self-Assessment Scores

Both correlations between the immersion SOPA ratings and SSA total scores and comparisons of immersion SOPA ratings and corresponding mean total SSA scores gave only partial support to the hypothesis that the higher the immersion SOPA rating, the higher the total score on the SSA should be. Correlations were neither high nor significant for either subset F or G (total immersion and two-way partial immersion).

The lack of support for hypotheses concerning the immersion SOPA and the SSA may be due to two factors: 1) differences in the way the two instruments assess proficiency and 2) differences in how they are rated. First, the two instruments are different in that the SSA asks students to silently reflect while the SOPA asks students to actively demonstrate their proficiency. The SSA consists of a number of written statements that the student rates his/her performance against, whereas the SOPA engages students in a series of performance tasks.

Secondly, rating differences involve not only the rating scale used but also what students are rated on and who assigns the rating. The SOPA rates students on a 9-point scale for four different skill areas whereas the SSA uses a 3-point scale for statements that touch upon three different skill areas (listening comprehension, speaking, and vocabulary). In addition, all students were rated by the same experienced CAL rater on the SOPA, whereas the SSA asked students to rate their own proficiency in the target language. Differences in self-perception and confidence level and attitudes towards the target language and culture are just a few of the variables that may have influenced students' perception of their own language ability.

SOPA Levels and Teachers' FLOSEM Ratings

Correlations and comparisons of means between SOPA levels and FLOSEM ratings for subsets F and G (total immersion and two-way partial immersion) provide evidence for the validity of the SOPA as an oral language skills assessment. The stronger correlations provided by the two-way immersion data may be due to a higher level of teacher training and experience with observation scales. Teachers have been using similar assessments for a number of years at the two-way immersion site, whereas teachers have not at the total immersion site. In addition, the range of scores provided by the two-way immersion data were broader and more varied than the total immersion data. Thus, the weaker correlations may also be a function of the homogenous nature of the total immersion data set.

VIII. Methodology – Study 3: Inter-Rater Reliability

The opportunity to conduct a small-scale, preliminary inter-rater reliability study became available because of the particular circumstances at one of the FLES sites in Study 1. At this site, students were not only rated by an experienced CAL rater but also by three local foreign language educators (Raters 1, 2, and 3) who had expressed interest in gaining experience in administering and rating the SOPA. Raters 1 and 3 had had no prior exposure to the SOPA or the SOPA rating scale. Rater 2 had attended a 2-day familiarization workshop on the SOPA, where the emphasis had been primarily on administration rather than rating. Although this study was not conducted with the typical rigor of an inter-rater reliability study where all raters would have participated in the requisite training, it was conducted because the circumstances were intriguing and it was hoped that the results would inform future reliability studies.

On the first day of the site visit, two of the foreign language educators (Raters 1 and 2) were present. On-site training consisted of a brief discussion of the SOPA and rating scale and then a demonstration during which the CAL rater interviewed and rated a pair of students. A brief discussion followed about interviewing techniques and how to interpret and assign ratings. The two "raters in training" then rated a number of interviews. They also took turns administering the SOPA. On the second day of SOPA administration, the third educator (Rater 3) gained experience administering the SOPA as well as rating interviews. This rater had received some orientation to the SOPA from the raters who had been present on the first day of the site visit. The CAL rater was not on-site the second day, but rated video-tapes of all the SOPA interviews (n = 15) a few days later.

At the conclusion of the SOPA administration, all rating sheets were collected and returned to CAL along with videotapes of the SOPA interviews. For the SOPA validation study, only the CAL rater's ratings were used. The ratings assigned by Raters 1, 2, and 3 were used for the inter-rater reliability study (Study 3). Ratings for Raters 1, 2, and 3 were compiled and correlated with the CAL rater's ratings to see if, even under the minimal training circumstances described above, inter-rater reliability could be established.

For this study, the data were examined in two ways: first in terms of interviews that were rated by three out of four raters (correlations across common examinees) and then in terms of all interviews that were rated by a least two raters (correlations between pairs of raters). The results of these analyses, as well as conclusions and recommendations, are presented below.

IX. Results of Data Analysis for Inter-Rater Reliability Study

Results of Correlations Across Common Examinees

Examination of the data revealed that a small number (10) of SOPA interviews had been rated by the same three raters (the CAL rater and Raters 1 and 2). Table 38 presents the correlations across common examinees. Across pairs of raters (CAL Rater and Rater 1, CAL Rater and Rater 2, and Raters 1 and 2), the average correlations for listening comprehension ratings (.77) and fluency ratings (.84) were moderate. When listening comprehension ratings are compared, the correlation between Raters 1 and 2 was the highest: .96 versus .65 for the correlation between the CAL rater and Rater 1 and .71 for the correlation between the CAL rater and Rater 2. For fluency ratings, the correlation between the CAL rater and Rater 2 was the highest (.92), followed by the correlation between the ratings by the CAL rater and Rater 1 (.82), and the correlation between Raters 1 and 2 (.79). The overall average correlations for all three raters were .77 for listening comprehension and .84 for fluency.

Table 38: Correlations Across Common Examinees (N = 10)

Comprehension	Rater 1	Rater 2		Average Correlation for all 3 Pairs of Raters
CAL Rater	.65	.71		.77
Rater 1		.96		
Fluency	Rater 1	Rater 2		Average Correlation for all 3 Pairs of Raters
CAL Rater	.82	.92		.84
Rater 1		.79		

Results of Correlations Between Pairs of Raters

Examination of the correlations between ratings assigned by all four raters revealed similar trends to the correlations across common examinees. Results are presented in Table 39 below. First comparisons involved the CAL rater and Raters 1, 2 and 3. Rater 1 and the CAL rater rated the same 49 students. The correlation between Rater 1 and the CAL rater was .50 for listening comprehension and .74 for fluency. Rater 2 and the CAL rater rated 15 students in common. Rater 3 and the CAL rater rated 12 students in common. Listening comprehension correlations were higher between Rater 2 and the CAL rater (.75) followed by .55 for Rater 3 and the CAL rater. For both pairings, the correlation for fluency was .84. The overall average (all pairs combined) correlation for listening comprehension was .60 and for fluency was .81.

Next, correlations between the local raters were examined. Raters 2 and 3 did not rate any of the same students so could not be compared. However, correlations between Raters 1 and 2 and Raters 1 and 3, who had rated some of the same students, were examined. Rater 1 and Rater 2 had a total of 10 students in common. The correlation between ratings for this pair of raters was .96 for listening comprehension and .79 for fluency. Rater 1 and Rater 3 rated the same 9 students. The correlation was .89 for listening comprehension and .66 for fluency. The average correlations between pairs of raters for listening comprehension ratings (.93) and fluency ratings (.73) were moderate to very good.

Table 39: Correlations Between Pairs of SOPA Raters

Comprehension	Rater 1	Rater 2	Rater 3		Average Correlation with Raters 1, 2, 3
CAL Rater	.50 (N = 49)	.75 (N = 15)	.55 (N = 12)		.60
Fluency	Rater 1	Rater 2	Rater 3		Average Correlation with Raters 1, 2, 3
CAL Rater	.74 (N = 49)	.84 (N = 15)	.84 (N = 12)		.81
Comprehension		Rater 2	Rater 3		Average Correlation
Rater 1		.96 (N = 10)	.89 (N = 9)		.93
Fluency		Rater 2	Rater 3		Average Correlation
Rater 1		.79 (N = 10)	.66 (N = 9)		.73

Summary of Results

Overall, correlations between the ratings assigned by the CAL rater and Raters 1, 2, and 3 were moderate. Higher correlations were not expected because of two factors. First, Raters 1, 2, and 3 were not fully-trained raters—they were given minimal training in the rating scale. Secondly, the CAL rater rated students (n = 15) off-site (via video) on the second day of SOPA administration, and consequently may not have experienced the SOPA interview in the same way as the on-site raters.

Conclusions and Recommendations

This preliminary study points to the importance of carefully controlling all circumstances when trying to establish inter-rater reliability. With revision of training procedures and materials, the consistency of ratings should increase. In addition, ensuring that interviews are all rated by trained raters under the same conditions would increase the

probability that each rater would observe the same behaviors and therefore assign similar ratings. Finally, the difference in correlations between on-site and on-site/off-site ratings touches on the intriguing question of what information contributes to the rating of listening comprehension, which unlike fluency, is inferred. The results of this preliminary study suggest that the cues and interactions that tell how well someone understands may not have been conveyed or interpreted the same way when video-taped.

X. Summary, Conclusions, and Future Directions

Summary

The goal of this research study was to assess the validity of the Student Oral Proficiency Assessment (SOPA), one of the few oral proficiency instruments that is designed specifically for children in elementary school language programs.

The subsequent data analysis was a two-step process. First, reports were prepared for each participating site that included average performance of students on each instrument (SOPA, a student self-assessment, and a teacher observation matrix) and comparisons of results for similar programs (July 1998 through November 1998). Second, the data were analyzed in terms of language and program type (5 data sets). Two different validity studies, one on the FLES data sets and one on the immersion data sets, were undertaken. Comparisons of mean ratings on the three instruments and correlations between the SOPA ratings and SSA total scores and the SOPA and FLOSEM ratings were used to determine the validity of the SOPA.

Overall, the studies provided moderate to strong support for the validity of the SOPA's claim to assess listening comprehension and speaking ability in a second language for young learners across language. Relationships between ratings on the SOPA and ratings based on teachers' observations (FLOSEM) were strong in 2 of the 5 data sets (across two languages and contexts) and moderate in the rest. The 2 data sets with the most variation among students (content-enriched Chinese FLES with second through fifth graders and heritage language speakers, and Spanish two-way immersion with both heritage and non-heritage language speakers) provided stronger empirical correlations between the SOPA and FLOSEM ratings than the other sites, where there was less variation among the students. The relationships between ratings on the SOPA and student self-assessment (SSA) were moderate for 1 data set and weak or nonexistent for the rest. The student self-assessment proved to be a generally unreliable measure (empirical reliability was also low, in the .60s). It seemed a particularly problematic measure for students to use in its immersion format.

A third, small-scale inter-rater reliability study was also conducted. This study compared listening comprehension and fluency ratings assigned by a highly trained SOPA rater with ratings assigned by three raters in training. Correlations between the ratings assigned were moderate, indicating the need for more extensive training of raters.

Conclusions

These studies provide moderate to strong support for the validity of the SOPA's claim to assess listening comprehension and speaking ability in a second language for young learners. The degree of support shown for the validity of the SOPA was directly related to the degree of variation among students at each site.

Although the Student Self-Assessment (SSA) did not prove to be a highly reliable measure when compared to the SOPA, feedback on the instrument from teachers at the different sites was positive. Many felt that the SSA provided valuable insight into student self-perception of language ability and, thus, could inform instructional content and delivery. Others suggested that expanding the rating scale from a 3-point to a 4- or 5-point range would increase the accuracy of student ratings. On the immersion SSA, it was suggested that asking students to rate their ability in both languages at the same time (French and English or Spanish and English) was confusing and may have impacted the reliability of the measure. Finally, item analysis also revealed that some items in the SSA "perform better" (contribute better to the SOPA's reliability) than others.

Future Directions

The results from these three studies and suggestions and observations made during the field-testing of the SOPA at each of the sites provided clear indications of how the SOPA and SOPA administrator's manual should be fine tuned for greater ease of use and accuracy of ratings. Once revisions were made (2001), the SOPA and its related materials have been made available to interested foreign language educators. Given the great interest in and need for student self-assessment of language skills, CAL project staff also revised both forms of the SSA to more accurately capture students' self-perceptions of their skills.

Three issues arose from the validation study that need to be addressed in future studies of the SOPA. First and foremost, results indicated that more research needed to be conducted on the inter-rater reliability of the instrument. Since the completion of the SOPA validation study, information on inter-rater reliability has been gathered on the SOPA in conjunction with program evaluations for school districts in which trained teachers as well as CAL researchers have served as raters. These results need to be formally analyzed and reported. More knowledge about the accuracy and inter-rater reliability of raters will provide valuable support and increased credibility for assessment results in foreign language programs nationwide. This knowledge will also inform current SOPA rater-training practices.

Second, follow-up SOPA validation studies have been recommended that would include sites with a wide range of variation among students (heritage language speakers, non-heritage language speakers, multiple grade levels, multiple languages, etc.) As related earlier, the amount of support shown by the other instruments for the validity of the SOPA was directly related to the degree of variation of students' background and language proficiency at each site.

A follow-up study with these parameters, as well as the use of other instruments that purport to measure proficiency, will better be able to illustrate the range of the SOPA proficiency results.

Lastly, since the revision of the SSA in 2001, information concerning correlations or comparisons between SOPA ratings and SSA ratings have been reported in conjunction with program evaluations for school districts. These results need to be formally analyzed and published as a part of future SOPA validation studies.

References

American Council on the Teaching of Foreign Languages. (1986). *ACTFL proficiency guidelines.* Yonkers, NY: Author.

American Council on the Teaching of Foreign Languages. (1998). *ACTFL performance guidelines for K-12 learners.* Yonkers, NY: Author.

Boyson, B., Rhodes, N., & Thompson, L. (1998, May). The Student Oral Proficiency Assessment—SOPA. *The ACIE Newsletter, 1*(4), 9, 13.

Clementi, D., & Sandrock, P. (1994). Putting our proficiency orientation into practice through meaningful assessment. *Report of Central States Conference on the Teaching of Foreign Language,* 91-102. Lincolnwood, IL: National Textbook.

Donato, R. (1998). Assessing foreign language abilities of the early language learner. In M. Met (Ed.), *Critical issues in early second language learning: Building for our children's future* (pp. 169-197). Glenview, IL: Scott Foresman.

Donato, R., Antonek, J. L., & Tucker, G. R. (1994). A multiple perspective analysis of a Japanese FLES program. *Foreign Language Annals, 27*(3), 365-378.

Educational Testing Service. (1993). *The ETS test collection catalog: Vol. 1. Achievement tests and measurement devices* (2nd ed.). Phoenix, AZ: Oryx.

Genesee, F., & Upshur, J. (1996). *Classroom-based evaluation in second language education.* New York: Cambridge University Press.

Heining-Boynton, A. (1990). The development and testing of the FLES program evaluation inventory. *The Modern Language Journal, 74,* 432-439.

Lapkin, S., Argue, V., & Foley, K. (1992). Annotated list of French tests: 1991 update. *The Canadian Modern Language Review, 48*(4), 780-807.

Lipton, G. (Ed.). (1996). *Evaluating FLES* programs: National FLES* commission report.* Champaign, IL: American Association of Teachers of French.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13-103). NY: American Council on Education and Macmillan.

National Standards in Foreign Language Education Project. (1996). *Standards for foreign language learning: Preparing for the 21st century.* Lawrence, KS: Allen Press.

Rhodes, N. (1988). Assessment instruments for immersion students and programs. In C. A. Klee, A. Lynch, & E. Tarone (Eds.), *Research and practice in immersion education: Looking back and looking ahead. Selected conference proceedings* (pp. 117-132). (CARLA Working Paper Series No. 10). Minneapolis, MN: The Center for Advanced Research on Language Acquisition.

Rhodes, N. (1998). Alternative assessment for immersion students: The Student Oral Proficiency Assessment (SOPA). In J. Arnau & J. M. Artigal (Eds.), *Els programes d'immersio: una perspectiva Europea* [Immersion programmes: A European perspective] (pp. 718-730). Barcelona, Spain: Universitat de Barcelona.

Rhodes, N., Rosenbusch, M., & Thompson, L. (1997). Foreign languages: Instruments, techniques, and standards. In G. Phye (Ed.), *Handbook of classroom assessment* (pp. 381-415). San Diego, CA: Academic Press.

Rhodes, N., & Thompson, L. (1990). An oral assessment instrument for immersion students: COPE. In A. M. Padilla, H. H. Fairchild, & C. Valadez (Eds.), *Foreign language education: Issues and strategies* (pp. 75-94). Newbury Park, CA: Sage.

Teachers of English to Speakers of Other Languages. (2001). *Scenarios for ESL standards-based assessment.* Alexandria, VA: Author.

Thompson, L. (1997). *Foreign language assessment in grades K-8: An annotated bibliography of assessment instruments.* Washington, DC and McHenry, IL: Center for Applied Linguistics and Delta Systems.

Tucker, G. R., Donato, R., & Antonek, J. (1996). Documenting growth in a Japanese FLES program. *Foreign Language Annals, 29*(4), 539-550.

INSTRUMENTATION

Boyson, B., Rhodes, N., & Thompson, L. (1996). *Student Oral Proficiency Assessment (SOPA). FLES Version.* (Available from Center for Applied Linguistics, 4646 40th Street NW, Washington, DC 20016)

Padilla, A. (1994). *Stanford Foreign Language Oral Skills Evaluation Matrix (FLOSEM).* (Available from Dr. Amado Padilla, Stanford University, School of Education, Palo Alto, CA 94305)

Rhodes, N., & Thompson, L. (1996). *Student Oral Proficiency Assessment (SOPA). Immersion Version.* (Available from Center for Applied Linguistics, 4646 40th Street NW, Washington, DC 20016)

Thompson, L. (1996). *CAL Student Self-Assessment (SSA).* (Available from Center for Applied Linguistics, 4646 40th Street NW, Washington, DC 20016)

Appendix A

Center for Applied Linguistics
Student Oral Proficiency Assessment (SOPA)
Background Information Questionaire

A note to program administrators/foreign language coordinators/teachers: *The following information is necessary for the interpretation of interview (SOPA) results of students in your program. For official reports concerning the SOPA, only such data as necessary for proper interpretation of results will be used. Please complete the form and set aside time to discuss your responses with the Center for Applied Linguistics rater who will be coming to your school in May/June.*

Program Information

1. What **type of program** do you have?

____FLEX ____FLES ____Content-based FLES

____Immersion ____ Other (explain)

Please define:

_____

2. How many time a week do classes meet and for how long each time?

_____

3. When did the program begin? _____

4. Who/what initiated the program? _____

____Parents ____School board

____Teacher(s) ____Other (explain)

_____

5. How are students **selected** for your program?

____ Parental choice _____School imposed criteria (e.g. test scores, grades)

____ School requirement _____ Other (explain)

_____

6. What is/are the **target language(s)**?_____

7. Why was/were this/these **language(s) selected**?

____Local population ____Academic needs of students

____Status ____Existing teacher/staff resource

____Other (explain)

8. What is the **ethnic make-up** of the class/school? (please list percentages)

____African American ____Anglo ____Asian

____Hispanic ____Other (Specify:_____)

9. Are there any **native speakers** of the target language in the class/program/school? If yes, how many?

10. Rank the following areas in terms of the **emphasis** they are given in your program: (1 is highest; 6 is lowest)

____cross cultural understanding ____listening

____reading ____speaking

____writing ____other (explain)

11. Do you or does your program subscribe to a specific **methodology**? Please describe.

12. Do you follow a **curriculum**? _____ Can we have a copy?____

13. How was this curriculum developed?_____

_____

14. Do students receive content instruction in the target language? If yes, how many hours per week?

_____

Lang Arts	Math	Soc Studies	Science	Other*

_____

K

1

2

3

4

5

6

*extra-curricular activities: field trips, student exchanges, etc.

15. Is there anything else that you would like to tell us concerning your language program or your school in general?

16. Is there a **follow-up** language program after elementary school?

Staff Information

1. How many **teachers** are there in your program? _____

2. How many of these teachers are **native speakers** of the target language? _____

3. What are their **national origins**?

4. Has there been **staff continuity** in your program? _____

5. Is there anything else that you would like to tell us concerning the staff in your program?

_____

_____

_____

_____

Name(s) of person(s) completing this questionnaire:

_____

Title/position:_____

Work Address:_____

Telephone: _____Fax:_____

E-mail:_____

Name of school/district participating in this study:

_____

Please return this questionnaire to the CAL SOPA rater when she visits your school. Thank you.

Appendix B

CAL Student Self-Assessment for French
(FLES Version)

Name:_____School:_____Date:_____

***Instructions**: We would like you to help us find out how much students are learning and what kinds of things students are learning in the French program at your school. This activity is not for a grade. There are no right or wrong answers.*

Look at the sentences below. For each sentence indicate if the sentence describes what you know by circling "yes", "sort of", or "not yet." Choose "yes" if the sentence describes what you can do easily and comfortably in French. Choose "sort of" if the sentence describes what you are sometimes able to do with a little bit of difficulty. Choose "not yet" if the sentence describes something that you do not know or can not do yet.

1.	I can say hello in French and tell someone my name.	YES	SORT OF	NOT YET
2.	I can follow instructions in French like "Sit down," "Open your book," and "Touch your head."	YES	SORT OF	NOT YET
3.	I can **understand** the names for lots of things in French (for example, classroom objects, members of a family, colors, numbers).	YES	SORT OF	NOT YET
4.	I can **say** the names of lots of things in French (for example, classroom objects, members of a family colors, numbers).	YES	SORT OF	NOT YET
5.	I can make sentences in French like "I have a cat," or "The boy is reading a book."	YES	SORT OF	NOT YET
6.	I can look at a picture of everyday life (for example, a classroom or a playground) and say in French what is happening.	YES	SORT OF	NOT YET
7.	I can retell a story, such as a fairytale, in French.	YES	SORT OF	NOT YET
8.	I can give instructions in French like "Sit down" or "Touch your head."	YES	SORT OF	NOT YET
9.	I feel comfortable speaking in French.	YES	SORT OF	NOT YET
10.	I can talk about how I am feeling in French.	YES	SORT OF	NOT YET

Please answer the questions on the next page.

***Instructions**: Please read the following questions and write your answer. There are no right or wrong answers. Just tell us how you feel.*

1. What do you think you know best in French?

2. What do you think you still need to learn?

©CAL, 1998

CAL Student Self-Assessment for Language

(Two-Way Immersion Version)

Name:_____School:_____Date:_____

***Instructions**: We would like you to help us find out how much students are learning and what kinds of things students are learning in the immersion program at your school. This activity is not for a grade. There are no right or wrong answers.*

Look at the sentences below. For each sentence indicate if the sentence describes what you know by circling "yes", "sort of", or "not yet." Choose "yes" if the sentence describes what you can do easily and comfortably in English or Spanish. Choose "sort of" if the sentence describes what you are sometimes able to do with a little bit of difficulty. Choose "not yet" if the sentence describes something that you do not know or can not do yet.

1. I can say hello and tell someone my name:
   in **English** YES SORT OF NOT YET
   in **Spanish** YES SORT OF NOT YET

2. I can follow instructions like "Sit down," "Open your book," and "Touch your head."
   in **English** YES SORT OF NOT YET
   in **Spanish** YES SORT OF NOT YET

3. I can **understand** the names for lots of things (for example, classroom objects, members of a family, colors, numbers).
   in **English** YES SORT OF NOT YET
   in **Spanish** YES SORT OF NOT YET

4. I can **say** the names of lots of things (for example, classroom objects, members of a family colors, numbers).
   in **English** YES SORT OF NOT YET
   in **Spanish** YES SORT OF NOT YET

5. I can make sentences like "I have a cat," or "The boy is reading a book."
   in **English** YES SORT OF NOT YET
   in **Spanish** YES SORT OF NOT YET

6. I can look at a picture of a classroom or a playground and say what is happening.
   in **English** YES SORT OF NOT YET
   in **Spanish** YES SORT OF NOT YET

7. I can tell a story that I know, such as a fairytale.
   in **English** YES SORT OF NOT YET
   in **Spanish** YES SORT OF NOT YET

8. I can give instructions like "Sit down" or "Touch your head."
   in **English** YES SORT OF NOT YET
   in **Spanish** YES SORT OF NOT YET

9. I feel comfortable speaking...
   in **English** YES SORT OF NOT YET
   in **Spanish** YES SORT OF NOT YET

10. I can talk about how I am feeling...

in **English** YES SORT OF NOT YET
in **Spanish** YES SORT OF NOT YET

11. I can give my opinion and even convince others that I am right on issues that are important to me and other people such as school rules.
in **English** YES SORT OF NOT YET
in **Spanish** YES SORT OF NOT YET

12. I can talk about what I am studying in my classes (math, science, social studies, for example).
in **English** YES SORT OF NOT YET
in **Spanish** YES SORT OF NOT YET

13. I know how to speak politely when talking to adults like my teachers or the principal.
in **English** YES SORT OF NOT YET
in **Spanish** YES SORT OF NOT YET

14. When I listen to someone speaking, I understand everything.
in **English** YES SORT OF NOT YET
in **Spanish** YES SORT OF NOT YET

15. I speak very well.
in **English** YES SORT OF NOT YET
in **Spanish** YES SORT OF NOT YET

Instructions: Please read the following questions and write your answer. There are no right or wrong answers. Just tell us how you feel.

1. Are there things that you still need to learn how to say in **English**? YES NO If your answer is YES, please write some of the things you still need to learn to say:

_____

2. Are there things that you still need to learn how to say in **Spanish**? YES NO If your answer is YES, please write some of the things you still need to learn to say:

_____

3. Which language do you speak most often at home? English Spanish

Other _____

4. Does anybody else in your family speak Spanish? YES NO If your answer is YES, who?

_____

Thank you!

©CAL 1998

CAL Student Self Assessment for Language

(Two-Way Immersion Version)

Name:_____School:_____Date:_____

***Instructions**: We would like you to help us find out how much students are learning and what kinds of things students are learning in the immersion program at your school. This activity is not for a grade. There are no right or wrong answers.*

Look at the sentences below. For each sentence indicate if the sentence describes what you know by circling "yes", "sort of", or "not yet." Choose "yes" if the sentence describes what you can do easily and comfortably in English or Spanish. Choose "sort of" if the sentence describes what you are sometimes able to do with a little bit of difficulty. Choose "not yet" if the sentence describes something that you do not know or can not do yet.

1. I can say hello and tell someone my name:
   in **English** YES SORT OF NOT YET
   in **Spanish** YES SORT OF NOT YET

2. I can follow instructions like "Sit down," "Open your book," and "Touch your head."
   in **English** YES SORT OF NOT YET
   in **Spanish** YES SORT OF NOT YET

3. I can **understand** the names for lots of things (for example, classroom objects, members of a family, colors, numbers).
   in **English** YES SORT OF NOT YET
   in **Spanish** YES SORT OF NOT YET

4. I can **say** the names of lots of things (for example, classroom objects, members of a family colors, numbers).
   in **English** YES SORT OF NOT YET
   in **Spanish** YES SORT OF NOT YET

5. I can make sentences like "I have a cat," or "The boy is reading a book."
   in **English** YES SORT OF NOT YET
   in **Spanish** YES SORT OF NOT YET

6. I can look at a picture of a classroom or a playground and say what is happening.
   in **English** YES SORT OF NOT YET
   in **Spanish** YES SORT OF NOT YET

7. I can tell a story that I know, such as a fairytale.
   in **English** YES SORT OF NOT YET
   in **Spanish** YES SORT OF NOT YET

8. I can give instructions like "Sit down" or "Touch your head."
   in **English** YES SORT OF NOT YET
   in **Spanish** YES SORT OF NOT YET

9. I feel comfortable speaking...
   in **English** YES SORT OF NOT YET
   in **Spanish** YES SORT OF NOT YET

10. I can talk about how I am feeling...

in **English** YES SORT OF NOT YET
in **Spanish** YES SORT OF NOT YET

11. I can give my opinion and even convince others that I am right on issues that are important to me and other people such as school rules.
in **English** YES SORT OF NOT YET
in **Spanish** YES SORT OF NOT YET

12. I can talk about what I am studying in my classes (math, science, social studies, for example).
in **English** YES SORT OF NOT YET
in **Spanish** YES SORT OF NOT YET

13. I know how to speak politely when talking to adults like my teachers or the principal.
in **English** YES SORT OF NOT YET
in **Spanish** YES SORT OF NOT YET

14. When I listen to someone speaking, I understand everything.
in **English** YES SORT OF NOT YET
in **Spanish** YES SORT OF NOT YET

15. I speak very well.
in **English** YES SORT OF NOT YET
in **Spanish** YES SORT OF NOT YET

Instructions: Please read the following questions and write your answer. There are no right or wrong answers. Just tell us how you feel.

1. Are there things that you still need to learn how to say in **English**? YES NO If your answer is YES, please write some of the things you still need to learn to say:

_____

2. Are there things that you still need to learn how to say in **Spanish**? YES NO If your answer is YES, please write some of the things you still need to learn to say:

_____

3. Which language do you speak most often at home? English Spanish

Other _____

4. Does anybody else in your family speak Spanish? YES NO If your answer is YES, who?

_____

Thank you!

©CAL 1998